

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Point of Contact:
 Jan Delaval
 Librarian-Physical Sciences
 CM1 1E01 Tel: 308-4498

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>Jan</u>	NA Sequence (#) <u>✓</u>	STN _____
Searcher Phone #: <u>4498</u>	AA Sequence (#) <u>✓</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit: _____
Date Searcher Picked Up: <u>9/14</u>	Bibliographic _____	Dr. Link _____
Date Completed: <u>9/12/01</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>✓</u>
Clerical Prep Time: <u>10</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>15</u>	Other _____	Other (specify) _____

This Page Blank (copy)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2001, 11:38:28 : Search time 2570.69 Seconds
(without alignments)
11709.012 Million cell updates/sec

Title: US-09-192-611-1
Perfect score: 1946
Sequence: 1 ACAGTGTGGAGATGGCGA.....CGCGATTCGCGCCGACCTGAG 1946

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 134457 seqs, 773874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_in4:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_rod:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1946	100.0	1946	9 AR028527	AR028527 Sequence
2	1946	100.0	1946	94 MM076759	U76759 Mus musculu
3	741.4	38.1	2576	9 AX004160	AX004160 Sequence
4	721.6	37.1	1260	9 AX004161	AX004161 Sequence
5	721.6	37.1	3047	85 AB050511	AB050511 Macaca fa
6	80	4.1	206008	76 AC084162	AC084162 Mus muscu
7	80	4.1	210695	76 AC083753	AC083753 Mus muscu
8	54.2	2.8	173126	86 AC005988	AC005988 Homo sapi

9	51	2.6	224583	75	AC074153	AC074153 Mus muscu
10	49.8	2.6	149719	78	AC090977	AC090977 Mus muscu
11	49.2	2.5	214574	65	AC019272	AC019272 Mus muscu
12	48.2	2.5	11696	3	SCPSCL	X54107 S.clavulige
13	47.4	2.4	160893	74	AC069456	AC069456 Homo sapi
14	47.2	2.4	77836	67	AC022573	AC022573 Homo sapi
15	47	2.4	147225	67	AC022498	AC022498 Homo sapi
16	46.4	2.4	5123	59	HEPVIE	X15120 Pseudorabie
17	46.4	2.4	8438	59	SHILTR	M57505 Pseudorabie
18	46	2.4	482	7	AF251348	AF251348 Canis fam
19	46	2.4	249687	63	AC015693	AC015693 Homo sapi
20	45.8	2.4	34862	3	SCG20A	AL360055 Streptomy
21	45.6	2.3	173561	64	AC016240	AC016240 Homo sapi
22	45.4	2.3	166341	63	AC013381	AC013381 Homo sapi
23	45.2	2.3	139627	64	AC016779	AC016779 Oryza sat
24	45.2	2.3	160285	88	AC090945	AC090945 Homo sapi
25	45.2	2.3	162403	23	AC016922	AC016922 Homo sapi
26	45	2.3	533	54	G57145	G57145 SHCC-102750
27	44.8	2.3	123580	2	AF263912	AF263912 Streptomy
28	44.8	2.3	152114	83	AP003448	AP003448 Oryza sat
29	44.8	2.3	173154	76	AC079933	AC079933 Trypanoso
30	44.6	2.3	51914	73	AC068022	AC068022 Homo sapi
31	44	2.3	4242	9	AR106298	AR106298 Sequence
32	44	2.3	5222	93	H5FXDNA	X61378 H.sapiens f
33	44	2.3	168013	78	AF261720	AF261720 Homo sapi
34	44	2.3	209887	78	AC090649	AC090649 Mus muscu
35	43.4	2.2	10022	1	AE005873	AE005873 Caulobact
36	43.2	2.2	588	89	AF251350	AF251350 Macaca ar
37	43	2.2	122395	78	AC090433	AC090433 Chlamydom
38	42.8	2.2	225257	74	AC069214	AC069214 Homo sapi
39	42.6	2.2	144893	79	AL355486	AL355486 Homo sapi
40	42.6	2.2	162345	90	AL161793	AL161793 Human DNA
41	42.6	2.2	186200	81	AL512306	AL512306 Homo sapi
42	42.4	2.2	96299	86	AC006330	AC006330 Homo sapi
43	42.2	2.2	20102	3	SC9B2	AL035212 Streptomy
44	42.2	2.2	69173	69	AC025052	AC025052 Homo sapi
45	42	2.2	20622	94	AF230869	AF230869 Mus muscu

ALIGNMENTS

RESULT 1					
AR028527	AR028527	1946 bp	DNA	PAT	29-SEP-1999
LOCUS	Sequence 1	from patent US 5858711.			
DEFINITION	AR028527				
ACCESSION	AR028527.1	GI:5940500			
VERSION					
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1946)				
AUTHORS	Glincher, L.H. and Hodge, M.R.				
TITLE	NF-AT-Interacting protein NIP45 and methods of use therefor				
JOURNAL	Patent: US 5858711-A 1 12-JAN-1999;				
FEATURES	Location/Qualifiers				
source	1..1946				

BASE COUNT	438 a	500 c	562 g	446 t
ORIGIN				

Query Match 100.0% Score 1946: DB 9: Length 1946:
Best Local Similarity 100.0%: Pred. No. 0: Mismatches 0: Indels 0: Gaps 0:
Matches 1946: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY	1	ACAGTGTGGAGATGGCGGAACCACTGAGGGAGCTGGTCCGAGTCCCGGGTGGCCGA	60
DB	1	ACAGTGTGGAGATGGCGGAACCACTGAGGGAGCTGGTCCGAGTCCCGGGTGGCCGA	60
QY	61	GGCGCTGGAGAGCCCGAGGCGCCGCTGGCGGTCTCGCGCCGGCAGTCTCGGGCT	120
DB	61	GGCGCTGGAGAGCCCGAGGCGCCGCTGGCGGTCTCGCGCCGGCAGTCTCGGGCT	120

DB	61	GGCGCTGGAGAGCCCGAGGCGCCGCTGGCGGTCTCGCGCCGGCAGTCTCGGGCT	120
QY	121	AGGCTATTCCAGACACCGGCTTGTGACTGGTGCATGACAGCAGCAGAGAGTCTTG	180
DB	121	AGGCTATTCCAGACACCGGCTTGTGACTGGTGCATGACAGCAGCAGAGAGTCTTG	180
QY	181	GAAGTCGACAGCCAGTAGAGGTGCCGTGCGCCCTCCCGCGCGCTAAACCTGAG	240
DB	181	GAAGTCGACAGCCAGTAGAGGTGCCGTGCGCCCTCCCGCGCGCTAAACCTGAG	240
QY	241	CAGACAGGACAGTGCATGAAAGGGGGCGGAGGGGCTCGCGGACCCCGGTAC	300
DB	241	CAGACAGGACAGTGCATGAAAGGGGGCGGAGGGGCTCGCGGACCCCGGTAC	300
QY	301	TTGGTGCAGACGGCGGGCGGGCTGTGATCCGAGAGAGCGCGGTGTCCAGTG	360
DB	301	TTGGTGCAGACGGCGGGCGGGCTGTGATCCGAGAGAGCGCGGTGTCCAGTG	360
QY	361	TACTCCGGAGAGGTACAGAGACGCTCAACTCATTCCAGATTAATTCATTCCTTG	420
DB	361	TACTCCGGAGAGGTACAGAGACGCTCAACTCATTCCAGATTAATTCATTCCTTG	420
QY	421	CTGTGCTTCAAGGCTGAGAGATGAGGAGATCTGCAAAATTCGGAGTTCCCTCT	480
DB	421	CTGTGCTTCAAGGCTGAGAGATGAGGAGATCTGCAAAATTCGGAGTTCCCTCT	480
QY	481	GAGAGATGATCCCTGCTTCAGGTTCTCCCTGGAGAAAGCTCAGAAAGAGTGTAG	540
DB	481	GAGAGATGATCCCTGCTTCAGGTTCTCCCTGGAGAAAGCTCAGAAAGAGTGTAG	540
QY	541	AAAGAGAAAGAAATGGAAGATTTCGGACACAGACATCTCTCTTGGCCCAACT	600
DB	541	AAAGAGAAAGAAATGGAAGATTTCGGACACAGACATCTCTCTTGGCCCAACT	600
QY	601	TGCTCAAGGAAAGAAAGCAATACGAGAGGCTCCAGAGTAAAGGGAAGGAAC	660
DB	601	TGCTCAAGGAAAGAAAGCAATACGAGAGGCTCCAGAGTAAAGGGAAGGAAC	660
QY	661	AAGCGTCTCAAGATCTCGCTCTGAGGCCCAAGAGCAGCAGATCCAGCCCTT	720
DB	661	AAGCGTCTCAAGATCTCGCTCTGAGGCCCAAGAGCAGCAGATCCAGCCCTT	720
QY	721	CAGAGCAGAGATGAGTGTCTAGTGAAGGCTCTTGGCCACAGACTCTCGA	780
DB	721	CAGAGCAGAGATGAGTGTCTAGTGAAGGCTCTTGGCCACAGACTCTCGA	780
QY	781	CTTTTACATCAAGATCGGTGCGGGCTGACCTAGTAGACTGCTGACAGATGTG	840
DB	781	CTTTTACATCAAGATCGGTGCGGGCTGACCTAGTAGACTGCTGACAGATGTG	840
QY	841	GAGCCCTTCAGAAATGTGGATGATCAATGCGCAATCATTTGGGGTGTCTCAACAG	900
DB	841	GAGCCCTTCAGAAATGTGGATGATCAATGCGCAATCATTTGGGGTGTCTCAACAG	900
QY	901	ATTCTTTTGGTTTTGAGAGACTGAATCTCTACTACCCCTAGTACCTTAAG	960
DB	901	ATTCTTTTGGTTTTGAGAGACTGAATCTCTACTACCCCTAGTACCTTAAG	960
QY	961	CTTGAGATGGCTACATCATTTGTTGTGCTGTGCAAGCTCTTCAGAGGCGCACAGAG	1020
DB	961	CTTGAGATGGCTACATCATTTGTTGTGCTGTGCAAGCTCTTCAGAGGCGCACAGAG	1020
QY	1021	ACATCCAGAGCTCGGCTCCGGGTGCAAGGAGAGAAACACCAAGATGTTGAGATC	1080
DB	1021	ACATCCAGAGCTCGGCTCCGGGTGCAAGGAGAGAAACACCAAGATGTTGAGATC	1080
QY	1081	TCACTGTCTCTGATTTCTCTTAAAGTTCTCATGTCACTATGAGGAAGCATGGA	1140
DB	1081	TCACTGTCTCTGATTTCTCTTAAAGTTCTCATGTCACTATGAGGAAGCATGGA	1140
QY	1141	CTCTGTGACACAAAGCTCTCTTCTTTGATGGACAAAGTTTCAGGCAAGAGCTG	1200
DB	1141	CTCTGTGACACAAAGCTCTCTTCTTTGATGGACAAAGTTTCAGGCAAGAGCTG	1200

QY	1201	CCACCTATCTGGGGCCCGGAATCCGAGATCTCATCGAAGTCTGGGGCCGGAAGCTCTAC	1260
Db	1201	CCACTATCTGGGCTGGATCCGAGATCTCATCGAAGTCTGGGGCCGGAAGCTCTAC	1260
QY	1261	CCCTGTTCGAGCGCAAAAGCCCAAGACATGGAGACATAGTCCCAATTTTATTTATTTGATTT	1320
Db	1261	CCCTGTTCGAGCGCAAAAGCCCAAGACATGGAGACATAGTCCCAATTTTATTTATTTGATTT	1320
QY	1321	TTTTCGCCCCCAATAGGGCTTAACAGAACTGATTTAGAACTTTGTTACTTATTTATTTCTGG	1380
Db	1321	TTTTGCCCCCAATAGGGGCTTAACAGAACTGATTTAGAACTTTGTTACTTATTTATTTCTGG	1380
QY	1381	TGCTGGGGATTTGAACCCCAAGCATATGCATATGCTTAAGATGTATGAAGTGAAGCGCAAAAC	1440
Db	1381	TGCTGGGGATTTGAACCCCAAGCATATGCATATGCTTAAGATGTATGAAGTGAAGCGCAAAAC	1440
QY	1441	CAAGGCAATTAACCTTTTACGACCCCTCAGTAGAGCTGATAGTGCACAGATGACTCTTGG	1500
Db	1441	CAAGGCAATTAACCTTTTACGACCCCTCAGTAGAGCTGATAGTGCACAGATGACTCTTGG	1500
QY	1501	TAGTGTGTGTGGCTCTGTGTATGTTTGTGCTGTATTTTGGCAGCCCTGGGGCACAATAGAG	1560
Db	1501	TAGTGTGTGTGGCTCTGTGTATGTTTGTGCTGTATTTTGGCAGCCCTGGGGCACAATAGAG	1560
QY	1561	GGACCTTGGGCTCCCTACCATTTACAGTTCGCGTGGGCCCTTCCCTTCATCATGATGACTT	1620
Db	1561	GGACCTTGGGCTCCCTACCATTTACAGTTCGCGTGGGCCCTTCCCTTCATCATGATGACTT	1620
QY	1621	CTGTGAAGCTGACCTAATGTATAGTGTGTTGAACATAAATGAGTCTGCTTTGGGTGCACAG	1680
Db	1621	CTGTGAAGCTGACCTAATGTATAGTGTGTTGAACATAAATGAGTCTGCTTTGGGTGCACAG	1680
QY	1681	CTGTGGGCTTTGTGCCGCAAGTTGGAGCCAGCAGTGCATCTGACTTGGGACGTAGAGAA	1740
Db	1681	CTGTGGGCTTTGTGCCGCAAGTTGGAGCCAGCAGTGCATCTGACTTGGGACGTAGAGAA	1740
QY	1741	TGCATTTCTCTGTGTGGAGACACTGGGGTGCAGAAATATTAACAGAGTGCATACATGCTGTG	1800
Db	1741	TGCATTTCTCTGTGTGGAGACACTGGGGTGCAGAAATATTAACAGAGTGCATACATGCTGTG	1800
QY	1801	AAGCTGAGGACTAGGTGAAGATTAACGAGCGTTCGATTTTCAGCCCTGGGCTTCCTCTCT	1860
Db	1801	AAGCTGAGGACTAGGTGAAGATTAACGAGCGTTCGATTTTCAGCCCTGGGCTTCCTCTCT	1860
QY	1861	GCTTCACGAGACTCTAGCCAGTGTCTGTACACACTTCTTGGCATGAGACACCTAGTGTGA	1920
Db	1861	GCTTCACGAGACTCTAGCCAGTGTCTGTACACACTTCTTGGCATGAGACACCTAGTGTGA	1920
QY	1921	CGCGGGCGCGAATTCGGCCGACTCGAG	1946
Db	1921	CGCGGGCGCGAATTCGGCCGACTCGAG	1946

Query Match	Best Local Similarity	Score 1946;	DB 94;	Length 1946;
Matches 1946;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
1	ACAGTGTGGAGATGTGGCGGAACCACTAGAGGGGAGCTGTCCGAGAGTCCCGGGGTGGCGGA	60		
1	ACAGTGTGGAGATGTGGCGGAACCACTAGAGGGGAGCTGTCCGAGAGTCCCGGGGTGGCGGA	60		
61	GGCGCTGGAGAGCGCCGAGGGCGCCGTCGCGCGGTCCTCCGCGCCGCGAGCTTCGCGCT	120		
61	GGCGCTGGAGAGCGCCGAGGGCGCCGTCGCGCGGTCCTCCGCGCCGCGAGCTTCGCGCT	120		
121	AGGCTCATTTCCAGACACCGTCTGTGTGGACTTGTGATGACAGACGACGAGAGAGTCTTG	180		
121	AGGCTCATTTCCAGACACCGTCTGTGTGGACTTGTGATGACAGACGACGAGAGAGTCTTG	180		
181	GAGTTCGCGAGACCCAGTGAAGTGTGCGCGCTCCCGCGCTCCCGCGCGCTTAAACTGTAG	240		
181	GAGTTCGCGAGACCCAGTGAAGTGTGCGCGCTCCCGCGCTCCCGCGCGCTTAAACTGTAG	240		
241	CAGGACAGGACAGTGAAGTGAAGGGCGCCGAGGGGCGCTGGGGAGCGCCCGCGTACA	300		
241	CAGGACAGGACAGTGAAGTGAAGGGCGCCGAGGGGCGCTGGGGAGCGCCCGCGTACA	300		
301	TTGGTGCAGAGCGCGCGCGCGCTGTGATGCCGAGAGGCGCGGTGCTCCAGTG	360		
301	TTGGTGCAGAGCGCGCGCGCGCTGTGATGCCGAGAGGCGCGGTGCTCCAGTG	360		
361	TACTTCGCGGAGGTACAGAGCAGGCTCAACCTCATTTCCAGATTAATTCCTCTTGA	420		
361	TACTTCGCGGAGGTACAGAGCAGGCTCAACCTCATTTCCAGATTAATTCCTCTTGA	420		
421	CTGTGCTTCAGAGCTTGAAGATGAGGCGAGATCTGACAAATTTGCGAGTTCCTCT	480		
421	CTGTGCTTCAGAGCTTGAAGATGAGGCGAGATCTGACAAATTTGCGAGTTCCTCT	480		
481	GAGGATGATGCGCGCTTCAGGTTTCCCTCGGAGAAAGAGTCAAGAAAGAACTGTAG	540		
481	GAGGATGATGCGCGCTTCAGGTTTCCCTCGGAGAAAGAGTCAAGAAAGAACTGTAG	540		
541	AAAGAGAAAGAAATGAGAGATTTCCGAGACGAGCATCTCTCTTCCCAACT	600		
541	AAAGAGAAAGAAATGAGAGATTTCCGAGACGAGCATCTCTCTTCCCAACT	600		
601	TGCTCAAGGAACAAGCAGAAAGCATACGAGGCGCTCCAGAGCTTAGGAGTGAAC	660		
601	TGCTCAAGGAACAAGCAGAAAGCATACGAGGCGCTCCAGAGCTTAGGAGTGAAC	660		

Dd	601	TCGTCAAGAAACAAACAGACAAAGACATACGAGAGCCCTCCACAGCTAAAGCACTGAAC	660
Qy	661	AAGGCTCTCAAGATCTCCGCTCTGCTGCTAGAGCCCAAGACAGACAGCTCCAGCCCTT	720
Dd	661	AAGGCTCTCAAGATCTCCGCTCTGCTGCTAGAGCCCAAGACAGACAGCTCCAGCCCTT	720
Qy	721	CAGAGCACAGATGTATGAGAGTGTCCTAGTGTGAAGAGGACGTGTGTGCACAGAGCTCTCGA	780
Dd	721	CAGAGCACAGATGTATGAGAGTGTCCTAGTGTGAAGAGGACGTGTGTGCACAGAGCTCTCGA	780
Qy	781	CTCTTTACACTCAAGATCCGGTGGCCGGGCTGACCTAGTAGAGACTGTCTGTGAGATGTGCG	840
Dd	781	CTCTTTACACTCAAGATCCGGTGGCCGGGCTGACCTAGTAGAGACTGTCTGTGAGATGTGCG	840
Qy	841	GAGGCCCTTCAGAAATGTGTGTGTGATCACATGTGGCCAAATCATCTTGGGTGTCTCCAACAGG	900
Dd	841	GAGGCCCTTCAGAAATGTGTGTGTGATCACATGTGGCCAAATCATCTTGGGTGTCTCCAACAGG	900
Qy	901	ATTCTTTGGCTTTTGTGAGAGAGAGAACTGTCTCTACTGTCCACCCCTGTACCCCTAAAG	960
Dd	901	ATTCTTTGGCTTTTGTGAGAGAGAGAACTGTCTCTACTGTCCACCCCTGTACCCCTAAAG	960
Qy	961	CTTGGAGTGCTGACATCATTTGATTTGTGTGTGCTAGCAAGCTCTTCAGAGGCCACAGAG	1020
Dd	961	CTTGGAGTGCTGACATCATTTGATTTGTGTGTGCTAGCAAGCTCTTCAGAGGCCACAGAG	1020
Qy	1021	ACATCCCAAGAGGCTCCGGGCTCCGGGTCAGAGGGAAGAGAAACACAGATGTGTGAGATTC	1080
Dd	1021	ACATCCCAAGAGGCTCCGGGCTCCGGGTCAGAGGGAAGAGAAACACAGATGTGTGAGATTC	1080
Qy	1081	TCACATGTCCTGATTTCTCTCTTAAAGGTTCTCATGTCACTATGAGAGAACCCATGTGGA	1140
Dd	1081	TCACATGTCCTGATTTCTCTCTTAAAGGTTCTCATGTCACTATGAGAGAACCCATGTGGA	1140
Qy	1141	CTCTCTGGACACAGCTCTCTCTTCTTGTGATGGACAAAGCTTTCAGGCAAGAGACTG	1200
Dd	1141	CTCTCTGGACACAGCTCTCTCTTCTTGTGATGGACAAAGCTTTCAGGCAAGAGACTG	1200
Qy	1201	CCAGCTGATCTGGGACCTGAGATCCGAGATCTCATGGAAGTGGGGCTGAAGCTCTCAC	1260
Dd	1201	CCAGCTGATCTGGGACCTGAGATCCGAGATCTCATGGAAGTGGGGCTGAAGCTCTCAC	1260
Qy	1261	CCTGTTGGGACGCAAGCTCTCTCTTCTTGTGAAGGGAACAAAGCTTTCAGGCAAGAGACTG	1320
Dd	1261	CCTGTTGGGACGCAAGCTCTCTCTTCTTGTGAAGGGAACAAAGCTTTCAGGCAAGAGACTG	1320
Qy	1321	TTTTGCCCCATTAAGGGCTTAACAGAAACTGAATTAGAACTTGTTACTTATTAATTTCTGAG	1380
Dd	1321	TTTTGCCCCATTAAGGGCTTAACAGAAACTGAATTAGAACTTGTTACTTATTAATTTCTGAG	1380
Qy	1381	TGCTGGGATTTGAACCCCAAGACTATGACATGCTAAGATGTATGAAGTGAAGGCCAAAC	1440
Dd	1381	TGCTGGGATTTGAACCCCAAGACTATGACATGCTAAGATGTATGAAGTGAAGGCCAAAC	1440
Qy	1441	CAAGGCATTACTTTAGCCAGCCCTTAATGATGTAGATGTAGTCAAGCAACTGGCTCTCTGG	1500
Dd	1441	CAAGGCATTACTTTAGCCAGCCCTTAATGATGTAGATGTAGTCAAGCAACTGGCTCTCTGG	1500
Qy	1501	TAGTTGTGTGCTGTGTATGTTTGTGTGTATTTGTGGAGGCCCTGTGGGACCATAGAG	1560
Dd	1501	TAGTTGTGTGCTGTGTATGTTTGTGTGTATTTGTGGAGGCCCTGTGGGACCATAGAG	1560
Qy	1561	GGACCTTGCGCTCCCTACCATTTACAGTTCGCTGGAGCCCTTCTCATCAAGATGACTT	1620
Dd	1561	GGACCTTGCGCTCCCTACCATTTACAGTTCGCTGGAGCCCTTCTCATCAAGATGACTT	1620
Qy	1621	CTGTGAAGCTCTCTATGTGTGATGTGTGTAACATAATGAGCTCTGCTTGGGTCACAG	1680
Dd	1621	CTGTGAAGCTCTCTATGTGTGATGTGTGTAACATAATGAGCTCTGCTTGGGTCACAG	1680
Qy	1681	CCTGGGCTTTTGTGCCAGATTTGGAAGCCAGCAGTGACTTCACTGTGACTTGGGACTGAGA	1740
Dd	1681	CCTGGGCTTTTGTGCCAGATTTGGAAGCCAGCAGTGACTTCACTGTGACTTGGGACTGAGA	1740

QY	1741	TGCATTTCCTG	TGGAGACACTCGGGTGCAGAAATATATACAGAAAGTACATAC	TGCTG	1800
Db	1741	TGCATTTCCTG	TGGAGACACTCGGGTGCAGAAATATATACAGAAAGTACATAC	TGCTG	1800
QY	1801	AAGCTGAGAGCTAG	TCGCAAACTTAACGACAGCTTGATTTTACGCTTGATATCC	TCTCT	1860
Db	1801	AAGCTGAGAGCTAG	TCGCAAACTTAACGACAGCTTGATTTTACGCTTGATATCC	TCTCT	1860
QY	1861	GCCTGCCAGAGCT	ATAGCCAGTGTCTGTACACACTTCTTGAGATGACACCTAGCTCGA	1920	
Db	1861	GCCTGCCAGAGCT	ATAGCCAGTGTCTGTGTGTACACACTTCTTGAGATGACACCTAGCTCGA	1920	
QY	1921	CGCGGGCGCGAT	TCGGCCGACTCGAG	1946	
Db	1921	CGCGGGCGCGAT	TCGGCCGACTCGAG	1946	
RESULT	3				
LOCUS	AX004160	2576 bp	DNA	PAT	24-AUG-2000
DEFINITION	Sequence 1 from Patent WO921993.				
ACCESSION	AX004160				
VERSION	AX004160.1	GI:9927712			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 2576)				
TITLE	Zhao,J. and Zhou,H.				
JOURNAL	Human nlp-45: Interleukin-4 gene transcriptional trans-activator and its uses				
FEATURES	Patent: WO 9921993-A 1 06-MAY-1999;				
source	ZENBICA LTD (GB)				
BASE COUNT	549 a 690 c 709 g 628 t				
ORIGIN					
Query Match	38.1%;	Score 741.4;	DB g;	Length 2576;	
Best Local Similarity	74.5%;	Pred. No. 9.4e-168;			
Matches 1033;	Conservative 0;	Mismatches 306;	Indels 48;	Gaps	6;
QY	6	GTGGGAGATGGGGGGAACCACTAGGGGAGCGTCCAGTCCCGCGGT-----GG	56		
Db	6	GTGGGAGATGGGGGGAACCACTAGGGGAGCGTCCAGTCCCGCGGT-----GG	56		
QY	57	CCGAGGCGCTCGAGAGGCCGAGGCCCGTGGCCGCTGCTCGCGCCGCGACATCTCC	116		
Db	57	CCGAGGCGCTCGAGAGGCCGAGGCCCGTGGCCGCTGCTCGCGCCGCGACATCTCC	116		
QY	66	CGGAGGGGGTGGGGGGGCGGTGGGGGCGTGGGGCGCGGTCTCGGGCCACGCGTCTCC	125		
Db	66	CGGAGGGGGTGGGGGGGCGGTGGGGGCGTGGGGCGCGGTCTCGGGCCACGCGTCTCC	125		
QY	117	GGCTAGGCTCATTTCCAGACACCGTCTTGTGTGACTTGTGTGACATGACACGAGAAAGT	176		
Db	117	GGCTAGGCTCATTTCCAGACACCGTCTTGTGTGACTTGTGTGACATGACACGAGAAAGT	176		
QY	126	ATTCGGGGGACAGCTGAGCTGTGTGTGTGACTTGTGTGACACCGACACGATGAGAAAT	185		
Db	126	ATTCGGGGGACAGCTGAGCTGTGTGTGTGACTTGTGTGACACCGACACGATGAGAAAT	185		
QY	177	CTTGGAGT-----GCGACACCACTAGAGTGCGCGTCCCGCGCT	218		
Db	177	CTTGGAGT-----GCGACACCACTAGAGTGCGCGTCCCGCGCT	218		
QY	186	TCTGGAGTGGCGCACCGGTCGCGCGCGCGGACGAGGTGAGTGGAGCCCGCGAGCC	245		
Db	186	TCTGGAGTGGCGCACCGGTCGCGCGCGCGGACGAGGTGAGTGGAGCCCGCGAGCC	245		
QY	219	CCCCGCGCGGCTAAACTGAGCAGAGACGACAGTATGACACTGAAGGGCGCGGAGG	278		
Db	219	CCCCGCGCGGCTAAACTGAGCAGAGACGACAGTATGACACTGAAGGGCGCGGAGG	278		
QY	246	CGCGGGCGCGGTGCGCTCCCGGATTAACGAAACAGTACACGCGAAGGAGGAGCGC	305		
Db	246	CGCGGGCGCGGTGCGCTCCCGGATTAACGAAACAGTACACGCGAAGGAGGAGCGC	305		
QY	279	GCCTCGGGAGCGCGCGACATTTGGTGGAGCGGGGGCGCGCGGTGTGATCCGG	338		
Db	279	GCCTCGGGAGCGCGCGACATTTGGTGGAGCGGGGGCGCGCGGTGTGATCCGG	338		
QY	306	GCCCGCAGAGACCCCGCGCGGAGCGGTACAGCGCGCGCGGTGTGATCCGG	365		
Db	306	GCCCGCAGAGACCCCGCGCGGAGCGGTACAGCGCGCGCGGTGTGATCCGG	365		
QY	339	AGAGCGCGCGGTGTCCAGTGTACTCCGGAAGGTACAGACAGCTCACTTATTC	398		
Db	339	AGAGCGCGCGGTGTCCAGTGTACTCCGGAAGGTACAGACAGCTCACTTATTC	398		

Db 366 GGAGGCGCCGCTGGTTCGGGTGTAAGTGAAGTTAAAGACGCTTCCCTTATCCC 425
QY 399 AGATTAATCATCCCTCTTGAACCTGTGCTTCAAGAGCTTGAAGATGAGGAGATCTGAC 458
Db 426 AGATGATCTATCCCTCCGGAAGCTCTACCTCCGAGGAGTGAAGAGGAGGAGGAGCTGGC 485
QY 459 AAATTTGGGAGTTCCTCTGAGAGTGAAGTCCCTGCTTCAAGTTCCTCCCTGAGAAA 518
Db 486 AGATTC---GAGTGGTCTCTACCATGAGGAGTCCCATCAGAGGCTCTCCCTGGAAGAC 542
QY 519 GAAGCTGAGAGAGAGTGAAG 578
Db 543 AAAGCTGAGAGAGTGAAG---GATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
QY 579 CATCTCTCTTTCGCCCCAACCTTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
Db 600 CAATCTCTCTCTGTCCTCCACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
QY 639 CCAGAGAGTGAAG 698
Db 660 CAAGAGAGTGAAG 719
QY 699 GCAG 758
Db 720 GCCACCTCAG 779
QY 759 TGTCTGGCAG 818
Db 780 CACCTTCCAG 839
QY 819 GAGAGTGCCTCTGAG 878
Db 840 CAGATGTCCTCTGAG 899
QY 879 TCTTGGGAGTCTCAAG 938
Db 900 CTTTGGGAGTCTCAAG 959
QY 939 TGCCACCCCTAGTACCTTAAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 998
Db 960 TGCCACCTCCAG 1019
QY 999 AAGCTCTTCAAG 1058
Db 1020 AAGTCTTCAAG 1079
QY 1059 GAACAG 1118
Db 1080 GAACAG 1139
QY 1119 ACAGTATGAG 1178
Db 1140 CCAGTATGAG 1199
QY 1179 AAAGCTTCAAG 1238
Db 1200 AAAGCTTCAAG 1259
QY 1239 AGCTGGGAG 1297
Db 1260 GGTCTGGGAG 1319
QY 1298 CTCCCAATTTATTAATTTTTCGCCCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
Db 1320 C-----TTTCCCTTTTTCGCCCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1365
QY 1358 CTGTGTTT 1364
Db 1366 CTATATCT 1372

RESULT 4
AX004161

LOCUS AX004161 1260 bp DNA
DEFINITION Sequence 2 from Patent W0921993.
ACCESSION AX004161
VERSION AX004161.1 GI:927713
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1260)
AUTHORS Zhao, J. and Zhou, H.
TITLE Human nlp-45: Interleukin-4 gene transcriptional trans-activator
and its uses
JOURNAL Patent: WO 9921993-A 2 06-May-1999;
ZENECA LTD (GB)
FEATURES
source location/Qualifiers
1..1260
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 269 a 364 c 407 g 220 t
ORIGIN
Query Match 37.1%; Score 721.6; DB 9; Length 1260;
Best Local Similarity 76.4%; Pred. No. 5.8e-163;
Matches 934; Conservative 0; Mismatches 264; Indels 24; Gaps 3;
QY 48 CCGCGGTGCGGAGGCGCTCGGAG 107
Db 45 CCGTCCCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 104
QY 108 GCAGTCTCGGCTAGGCTCATTCAGAGACCGAGTGTGTGAGACTTGTCAGTGAACAGGCA 167
Db 105 GCGGTCTCATTCGCGGAGGAG 164
QY 168 CGAAGAGTCTTGAAGT-----GCGAGCCAGTAGAGAGTCCCGGT 209
Db 165 TGAGGAATTTCTGAGAGTCCGACCGCTCGGAGTCCGCGAGAGAGAGAGAGAGAGAGAG 224
QY 210 CGCCGCGCTCCCGCGCGGCGCTAAACCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 269
Db 225 CCGGAGGCGCGCGGCGGCGGCTCCCGGAGTAACAGCAACAGTGAAGAGAGAGAGAGAG 284
QY 270 GAGCGAGGAGGCGGCGGAG 329
Db 285 GAGAGAGGAGGCGGAG 344
QY 330 GGATCCGAGAGAGGCGGAGTCCAGTGTACTCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 389
Db 345 GGATCCGAGAGAGGCGGAGTCCAGTGTACTCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 404
QY 390 CCTCATTCAGATTAATTAATCCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 449
Db 405 CTTATTCAGATTAATTAATCCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 464
QY 450 AGATTCAGAAATTTCTGAGAGTTCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 509
Db 465 AGAGCTGGCAGATTC---GAGTGTCTTACCATGAGAGAGAGAGAGAGAGAGAGAGAGAG 521
QY 510 CTGGAAGAGAGAGTCAAG 569
Db 522 CTGGAAGAGAGAGTCAAG 578
QY 570 GGAGCAGAGATTCCTCTTTCGCCCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 629
Db 579 GGATTCGAGAGAGTCTCTCTGTCCTCCACCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 638
QY 630 GGAGGCGCTCCAGAGAGTGAAG 689
Db 639 TCGGCACTCAAGAGAGTGAAG 698
QY 690 GAGCCCAAG 749
Db 690 GAGCCCAAG 749

	Db	699	GACCCCAAGCACCCTCGAGGGTCAAGAGCACACGAGGCCAAGAGATGAATGGCTTTGGT	758
	Oy	750	GGAAAGCCCTGTCTTGCCACAGAGCTTCGACTCTTTAACTACAATYCCGGTCCGGGC	809
	Db	759	GGAAAGGGCCACCCCTCCAGAGACCCCCCGACTCTTCCCACTCAAAAATCCGTTCCGGGC	818
	Oy	810	TGACCTAGTAGAGCTGCCGTGCAGATGTGCGACCCCCTTCAGAAATGTGGTGGATCATAT	869
	Db	819	TGACCTGGTCAGATTGGCCCTTCAGAGATGTGCGACCCCCTTCAGAGTGTGGTGGACCAAT	878
	Oy	870	GGCCATCATCTTGGGGTGTCTCCAACAGAGATTTCTTTGGTTTTGGAGAAGTGAAT	929
	Db	879	GGCCACCCACCTTGGGGGTGTCCCAAGAGAGATCTTTTGCTTTTGGAGAGACAGAGCT	938
	Oy	930	GTCCTCTCTGCGCACCCCTTAAGTACCCTTAAGCTTGAGATGGCTGACATCTATTGTGT	969
	Db	939	ATCAGCTATCTGCACTCCAGAGACCTTAAAGCTCGAGTGGATGGCTGACATTAATGACTGT	998
	Oy	990	GGTGTACCAAGCTCTTCAGAGGGCCACAGACATATCCACAGAGCTCCGGCTCCGGGTGCA	1049
	Db	999	GGTACTAACCAAGTTCTCCAGAGGCCACAGAGACGTCCCAACAGCTCCAGCTCCGGGTGCA	1058
	Oy	1050	GGGGAAGAGAAACACCCAGATGTGTGGAGATCTCACTGTCTTCATTTCTCTTAAGGT	1109
	Db	1059	GGCAAGAGAGAAACCCAGACACTGGAGATCTCACTGTCTCAAGATTTCCCTCTAAAGAC	1118
	Oy	1110	TCTCATGTACACTGTATGAGAGAGCATGGAGCTCTGAGACACAGCTCTCTCTCTT	1169
	Db	1119	CCTCATGTGCCACTGTAGAGAGGCATGGAGCTGTGCGAGACGATCTCTCTCTT	1178
	Oy	1170	TGATGGGACAAAGCTTTTAGCGCAGAGAGCTCCACAGCTGATCTGGGCTGGAAATCCGAGA	1229
	Db	1179	TGATGGGACAAAGCTTTTAGCGCAGAGAGCTCCACAGCTGATCTGGGCTGGAAATCCGAGA	1238
	Oy	1230	TCTCATCGAAGCTCTGGGCTGA 1251	
	Db	1239	CCTCATTTAGAGTGTCTGGGCTGA 1260	
RESULT	5			
LOCUS	ABO50511	3047 bp	mRNA	PRI 27-OCT-2000
DEFINITION	Macaca fascicularis brain cDNA, clone:nPa-16750.			
ACCESSION	ABO50511			
VERSION	ABO50511.1 GI:11041476			
KEYWORDS	f1s (full insert sequence).			
SOURCE	Macaca fascicularis adult male brain parietal lobe cDNA to mRNA, clone:jlb:macaque brain cDNA library Gnpa clone:nPa-16750.			
ORGANISM	Macaca fascicularis			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;			
	Cercopithecinae; Macaca.			
AUTHORS	Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,			
	Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.			
TITLE	Isolation of full-length cDNA clones from macaque brain cDNA libraries			
JOURNAL REFERENCE	Unpublished			
AUTHORS TITLE	2 (bases 1 to 3047)			
JOURNAL	Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.			
	Direct Submission			
	Submitted (25-OCT-2000) to the DDBJ/EMBL/Genbank databases.			
	Katsuyuki Hashimoto, National Institute of Infectious Diseases,			
	Division of Genetic Resources, 23-1, Toyama 1-chome, Shinjuku-			
	ku, Tokyo 162-8640, Japan (E-mail:khashienh.go.jp,			
	URL:http://www.nih.go.jp/yoken/genbank/,			
	Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)			
COMMENT	Lab host: TOP10			
	Vector: pME18S-F13 (Acc.No. AB009864)			
	R. Site1: DraIII (CACTGTGTC)			
	R. Site2: DraIII (CACCATGTG)			
	Description: 1st strand cDNA was primed with an oligo(dT) primer			
	[ATGTCGCCCTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized			

using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5 kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pEMV8-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' and primer (CGACCTGCAGCTCTCAAGCTCGG); 3' end primer (CGACCTGCAGCTCTCAAGCTCGG)).

[illegible]

```

Db      531 AAGAAAAAGACAGAGATTCTGATCTGGACAACTCTCTCTGTCCACCCTTCACCGAGG 550
Qy      610 AACAAAAGCAGAAAGATACGAGGCGCTCCAGAGCTAAGGAGGAGTGAACAGCGCTC 669
Db      591 ACCAAAAGCAGAAAGATACGAGGCGCTCCAGAGCTAAGGAGGAGTGAACAGCGCTC 650
Qy      670 CAAGATCTCCGCTCTGCTGAGCCCCCAAGCAGACAGCAGAGTCCAGCCCTTCAGACACA 729
Db      651 CAGAGATCTCCGCTCTGCTGAGCCCCCAAGCAGAGCTAAGGAGGAGTGAACAGCGCTC 710
Qy      730 GATGATGAGGTGCTCTAGTGAAGGCGCTGTCTTGGCCACAGAGCTCTGACTCTTTACA 789
Db      711 GAGGAGAGGAGTGTCTTGGTGAAGGCGCCACCTCCAGAGAGCCCGACCTCTTCCCG 770
Qy      790 CTCAGATCTCCGCTCTGCTGAGCCCCCAAGCAGAGCTAAGGAGGAGTGAACAGCGCTC 849
Db      771 CTCAGATCTCCGCTCTGCTGAGCCCCCAAGCAGAGCTAAGGAGGAGTGAACAGCGCTC 830
Qy      850 CAGAGATCTCCGCTCTGCTGAGCCCCCAAGCAGAGCTAAGGAGGAGTGAACAGCGCTC 909
Db      831 CAGAGATCTCCGCTCTGCTGAGCCCCCAAGCAGAGCTAAGGAGGAGTGAACAGCGCTC 890
Qy      910 CTTTGGAGAGAGTGAAGTGTCTCTACTGCGACCCCTAGTACCTTAAAGCTTGGAGTG 969
Db      891 CTTTGGAGAGAGTGAAGTGTCTCTACTGCGACCCCTAGTACCTTAAAGCTTGGAGTG 950
Qy      970 GCTGACATCATTTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1029
Db      951 GCTGACATCATTTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1010
Qy      1030 GAGCTCCGCGCTCGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1089
Db      1011 CAGCTCCAGCTCGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1070
Qy      1090 CCGATTTCTCTCTTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1149
Db      1071 CCGATTTCTCTCTTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1130
Qy      1150 CACAGCTCTCTCTTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1209
Db      1131 CCGAGCTCTCTCTTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1190
Qy      1210 CTGGGCTCTCTCTTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1268
Db      1191 CTGGGCTCTCTCTTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1250
Qy      1269 GACGCAAGCAGAAAGATGAGAGCATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1328
Db      1251 ATGGCCAGCTCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1296
Qy      1329 CATTAAGGCTAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1364
Db      1297 CATTAAGGCTAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1332

```

```

RESULT 6
AC084162 206008 bp DNA HTG 07-FEB-2001
LOCUS Mus musculus chromosome 5 clone RP23-11P12 strain C57BL6/J, WORKING
DEFINITION DRAFT SEQUENCE, 14 unordered pieces.
ACCESSION AC084162 GI:12704629
VERSION AC084162.2 GI:12704629
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 206008)
REFERENCE 1
AUTHORS Ayale, R., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Buffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,
Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q.,
Legsapi, R., Lim, M., Maduro, O.L., Maduro, V.B., Mastiello, C.,
Mastrian, S.D., McLoskey, J.C., McDowell, J., Pearson, R., Prasad, A.,
Shevchenko, Y., Snyder, B., Stantrijop, S., Thomas, J.W., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,
Wetherby, K.D., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 206008)
REFERENCE 2
AUTHORS Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2000) NIH Intramural Sequencing Center, 8717
Groveomt Circle, Gaithersburg, MD 20877, USA
On Feb 7, 2001 this sequence version replaced gi:10801352.
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@ngri.nih.gov
----- Project Information
Center project name: rs
Center clone name: 01P12
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 20092 bases at least Q40
Consensus quality: 20157 bases at least Q30
Consensus quality: 20200 bases at least Q20
Insert size: 20200; agarose-fp
Insert size: 204708; sum-of-ctrls
Quality coverage: 11.68x in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2503: contig of 2503 bp in length
2504 2603: gap of unknown length
2604 6058: contig of 3455 bp in length
6059 6158: gap of unknown length
6159 12381: contig of 6233 bp in length
12382 12481: gap of unknown length
12482 17111: contig of 4630 bp in length
17112 17211: gap of unknown length
17212 24466: contig of 7255 bp in length
24467 24566: gap of unknown length
24567 33228: contig of 8662 bp in length
33229 33328: gap of unknown length
33329 40894: contig of 7566 bp in length
40895 40995: gap of unknown length
40996 52203: contig of 11209 bp in length
52204 52303: gap of unknown length
52304 64564: contig of 12261 bp in length
64565 64665: gap of unknown length
64666 76971: contig of 12307 bp in length
76972 77071: gap of unknown length
77072 92317: contig of 15246 bp in length
92318 92417: gap of unknown length
92418 113190: contig of 20773 bp in length
113191 113290: gap of unknown length
113291 159328: contig of 46038 bp in length
159329 159428: gap of unknown length
159429 206008: contig of 46580 bp in length.
Location/Qualifiers
1..206008
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-11P12"

```

```

misc_feature      /clone_lib="RPCi mouse BAC library 23"
                  1..2503
misc_feature      /note="assembly_fragment"
                  2604..6058
misc_feature      /note="assembly_fragment"
                  6159..12381
misc_feature      /note="assembly_fragment"
                  12482..17111
misc_feature      /note="assembly_fragment"
                  clone_end:SP6
                  vector_side:left"
misc_feature      17212..24466
misc_feature      /note="assembly_fragment"
                  24567..33228
misc_feature      /note="assembly_fragment"
                  33329..40894
misc_feature      /note="assembly_fragment"
                  40995..52203
misc_feature      /note="assembly_fragment"
                  52304..64564
misc_feature      /note="assembly_fragment"
                  64665..76971
misc_feature      /note="assembly_fragment"
                  77072..92317
misc_feature      /note="assembly_fragment"
                  92418..113190
misc_feature      /note="assembly_fragment"
                  113291..159328
misc_feature      /note="assembly_fragment"
                  159429..206008
misc_feature      /note="assembly_fragment"
                  206009..51451
BASE COUNT      51690 a 49057 c 51451 g 52506 t 1304 others
ORIGIN
Query Match      4.13: Score 80; DB 76; Length 206008;
Best Local Similarity 73.13; Pred. No. 7.1e-09;
Matches 117; Conservative 0; Mismatches 40; Indels 3; Gaps 1;
QY 903 TCTTTGCTTTTGGAGAGTGAAGTGTCTTCCAGCCCTGATACCTTAAGCT 962
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156786 TTTTCTGTTTGGAGAGGCTGACTGTCCCTTGTGCCACCCCTGATTTCTTAACCT 156845
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 963 TGAAGTGCATCATTCATTTGTTGTGTGTAGCAAGCTCTTCAGAGCCACAGAGAC 1022
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156846 TGGCTAGCGGACATCATCTGTTTCTGCTAGTAAGGCTTCCAAAGCTACTGAGAC 156902
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1023 ATCCAGAGCTCCGGCTCCGGGTGCGAGGAGAGAGAGAA 1062
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156903 ATCCAGAGCTCCGGCTCCGGGTGCGAGGAGAGAGAGAA 156942
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

TITLE            NISC Comparative Sequencing Initiative
JOURNAL          Unpublished
REFERENCE        2 (bases 1 to 210695)
AUTHORS          Green, E.D.
TITLE            Direct Submission
JOURNAL          Submitted (30-SEP-2000) NIH Intramural Sequencing Center, 8717
                  Grovemont Circle, Gaithersburg, MD 20877, USA
                  On Feb 21, 2001 this sequence version replaced g1:10440605.
COMMENT          ----- Genome Center
                  Center: NIH Intramural Sequencing Center
                  Center code: NISC
                  Web site: http://www.nisc.nih.gov
                  Contact: nisc.moussenhgrl.nih.gov
                  Project Information
                  -----
                  Center project name: rt
                  Center clone name: 480K09
                  ----- Summary Statistics
                  Sequencing vector: plasmid; n/a; 100% of reads
                  Chemistry: Dye-terminator Big Dye; 100% of reads
                  Assembly program: Phrap; version 0.990319
                  Consensus quality: 208188 bases at least Q40
                  Consensus quality: 208941 bases at least Q30
                  Consensus quality: 209303 bases at least Q20
                  Insert size: 20300; agarose-fp
                  Insert size: 209995; sum-of-ctrls
                  Quality coverage: 11.96x in Q20 bases; agarose-fp
                  Quality coverage: 11.56x in Q20 bases; sum-of-ctrls
                  -----
                  * NOTE: This is a 'working draft' sequence. It currently
                  * consists of 8 contigs. The true order of the pieces
                  * is not known and their order in this sequence record is
                  * arbitrary. Gaps between the contigs are represented as
                  * runs of N, but the exact sizes of the gaps are unknown.
                  * This record will be updated with the finished sequence
                  * as soon as it is available and the accession number will
                  * be preserved.
                  *
                  * 2686: contig of 2686 bp in length
                  * 2687 2786: gap of unknown length
                  * 2787 8537: contig of 5751 bp in length
                  * 8538 8637: gap of unknown length
                  * 8638 17518: contig of 8881 bp in length
                  * 17519 17618: gap of unknown length
                  * 17619 33828: contig of 16210 bp in length
                  * 33829 33928: gap of unknown length
                  * 33929 74177: contig of 40248 bp in length
                  * 74177 74276: gap of unknown length
                  * 74277 114480: contig of 40204 bp in length
                  * 114481 114580: gap of unknown length
                  * 114581 151240: contig of 36650 bp in length
                  * 151241 151340: gap of unknown length
                  * 151341 210695: contig of 59355 bp in length.
                  *
                  Location/Qualifiers
                  1..210695
                   /organism="Mus musculus"
                   /strain="C57BL6/J"
                   /db_xref="taxon:10090"
                   /chromosome="5"
                   /clone_lib="RP23-480K9"
                   /clone_lib="RPCi mouse BAC library 23"
                   1..2686
                    /note="assembly_fragment"
                    2787..8537
                     /note="assembly_fragment"
                     8638..17518
                      /note="assembly_fragment"
                      clone_end:77
                      vector_side:right"
                      17619..33828
                       /note="assembly_fragment"
                       33929..74176
                        /note="assembly_fragment"
                        74277..114480
                         /note="assembly_fragment"

```

misc_feature	114581..151240	/note="assembly_fragment"
misc_feature	151341..210655	/note="assembly_fragment"
	clone_end:SP6	vector_side:left"
BASE COUNT	52952 a 51898 c 52183 g 52957 t 705 others	
ORIGIN		
Query Match	4.1%, Score 80; DB 76; Length 210695;	
Best Local Similarity	73.1%; Pred. No. 7.1e-09;	
Matches 117; Conservative	0; Mismatches 40; Indels 3; Gaps 1;	
Oy 903	TCCTTTGCTTTTGGAGAGAGTGAACCTGTCCTTCCTGACCCCTAGTACCTTAAGCT 962	
Db 160754	TTTTTCCTTTTGGAGAGGCTGAACCTGTCCTTCCTGACCCCTAGTATCTTAACCT 160695	
Oy 963	TGGAAGTGCATGACATCATTTGATTTGTTGTTGCTTACCAAGCTCTTTCAGAGCCACAGAGAC 1022	
Db 160694	TGGTTGATCCGACGACATCATCTGTTTG---CTGCTAAGTAAGAGCCTTCACAGGCTATCTAGAGAC 160638	
Oy 1023	ATCCAGAGAGCTCCGGCTCCGGGTCCAGAGGAGAGAA 1062	
Db 160637	ATCCAGAGAGTCCGACCTCCGGGAGAGAGAAACAA 160598	
RESULT 8		
LOCUS	AC005988 173126 bp DNA PRI 15-JAN-1999	
DEFINITION	Homo sapiens chromosome 17, clone hnpk.299_G_24, complete sequence	
ACCESSION	AC005988	
VERSION	AC005988.1 GI:4156128	
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 173126)	
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 173126)	
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donnell,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gerlsey,G., Grant,G., Hagos,B., Heaford,A., Herens,L., Horton,L., Howland,J.C., Jactot,L., Jones,C., Kann,L., Karats,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurt,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahl,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P., Strange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Yee,W.J., Zhao,J. and Zody,M.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (19-NOV-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
AUTHORS	3 (bases 1 to 173126)	
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeCarliano,K., Depayre,E., Devon,K., Dewar,K., Donnell,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karats,A., Lehoczy,J., Liu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurt,A., McKernan,K., McLaughlin,J., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,	

REFERENCE	AUTHORS
TITLE	Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
JOURNAL	Stange-Thomas,N., Stojanovic,N., Stone,C., Subramanian,A.,
	Testa,G.S., Toruella-Miller,T., Vassiliev,H., Vo,A., Wagner,A.,
	Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
	Direct Submission
	Submitted (14-JAN-1999) Whitehead Institute/MIT Center for Genome
	Research, 320 Charles Street, Cambridge, MA 02141, USA
	4 (bases 1 to 173126)
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.
	Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
	Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
	Cooke,P., DeRubeano,K., Depayre,E., Devon,K., Dewar,K.,
	Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
	Funke,R., Gage,D., Galagan,J., Gaidyna,S., Gilbert,D., Grant,G.,
	Hagos,B., Heisford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
	Karats,N., Lehoczy,J., Iken,C., Locke,K., MacDonald,P.,
	Marquis,N., McEwan,P., McGuck,A., McKernan,K., McLaughlin,J.,
	Meldrum,J., Molla,M., Morris,N., Morrow,J., Mychaleckyj,J.,
	Naylor,J., Nijhoff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
	Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
	Stange-Thomas,N., Stojanovic,N., Stone,C., Subramanian,A.,
	Testa,G.S., Toruella-Miller,T., Vassiliev,H., Vo,A., Wagner,A.,
	Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (15-JAN-1999) Whitehead Institute/MIT Center for Genome
	Research, 320 Charles Street, Cambridge, MA 02141, USA
	On Jan 14, 1999 this sequence version replaced g1:415385.
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. &
	Green, P. (1996-1997)
	http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only the first 173126 base pairs of this clone are being submitted
The remainder overlaps accession number AC005988 (WICGR project
L405).

FEATURES	source
Location/Qualifiers	1..173126
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="hRPK.299_G-24"
	/clone_id="RPCL11 human BAC library"
	/map="17"
	/chromosome="17"
repeat_region	464..760
	/rpt_family="AluCb"
repeat_region	904..1065
	/rpt_family="AluSg/x"
repeat_region	complement(1211..1279)
	/rpt_family="MIR"
repeat_region	6830..6858
	/rpt_family="TTTTTGGn"
repeat_region	7229..7251
	/rpt_family="(TTTA)n"
repeat_region	complement(7252..7533)
	/rpt_family="AluSg"
repeat_region	9423..9450
	/rpt_family="(T)n"
repeat_region	complement(9523..9632)
	/rpt_family="L2"
repeat_region	9946..10122
	/rpt_family="AluCb"
repeat_region	11028..11239
	/rpt_family="MER20"
repeat_region	11471..11519
	/rpt_family="(TC)n"
repeat_region	11519..11541
	/rpt_family="(TG)n"
repeat_region	11553..11678
	/rpt_family="L2"
repeat_region	12346..12447
	/rpt_family="(CA)n"
repeat_region	complement(12457..12613)
	/rpt_family="MIR"
repeat_region	complement(12829..12954)

```
repeat_region      /rpt_family="MIR"
complement(13723. .14086)
repeat_region      /rpt_family="THER1B"
complement(14191. 14483)
repeat_region      /rpt_family="AluSg"
complement(15046. .15333)
repeat_region      /rpt_family="AluY"
complement(15365. .15396)
repeat_region      /rpt_family="AluSx"
15397. 15420
/rpt_family="(TTTTC)n"
repeat_region      complement(15421. .15706)
/rpt_family="AluSx"
repeat_region      complement(15812. .15859)
/rpt_family="L2"
16016. 16462
/rpt_family="MSTB"
16463. .16739
repeat_region      /rpt_family="MSTB-internal"
16849. .17257
repeat_region      /rpt_family="MSTC-internal"
17258. 17552
/rpt_family="AluYB8"
17533. .17860
repeat_region      /rpt_family="MSTC-internal"
17861. 18280
/rpt_family="MSTC"
18733. .19010
repeat_region      /rpt_family="L1ME"
complement(19191. .19322)
repeat_region      /rpt_family="L1MD1"
complement(19331. .19383)
repeat_region      /rpt_family="AluSg/x"
19610. 19637
/rpt_family="(TG)n"
repeat_region      complement(19646. .19650)
/rpt_family="L1MD2"
20361. .20431
repeat_region      /rpt_family="MLT1L"
complement(21186. .21364)
/rpt_family="MER20"
21762. 21895
/rpt_family="MIR"
22333. .22490
repeat_region      /rpt_family="MER20"
22554. .22841
/rpt_family="AluYB"
repeat_region      complement(23485. .23708)
/rpt_family="MIR"
24098. 24138
/rpt_family="(TTTA)n"
repeat_region      complement(24139. .24830)
/rpt_family="L1PA3"
25054. .25385
/rpt_family="L1MB5"
25386. 25407
/rpt_family="(CA)n"
repeat_region      25408. .25770
/rpt_family="L1MB5"
26223. .26330
/rpt_family="L1MB3"
repeat_region      complement(26689. .26703)
/rpt_family="L1ME3"
26704. .26726
/rpt_family="(CA)n"
repeat_region      complement(26727. .26824)
/rpt_family="L1ME3"
complement(26825. .27191)
/rpt_family="THER1B"
repeat_region      complement(27192. .27348)
/rpt_family="L1ME3"
29352. 29487
/rpt_family="MIR"
```

```
repeat_region      29508. .29549
/rpt_family="AT_rich"
repeat_region      complement(29933. .30113)
/rpt_family="MLT1T"
repeat_region      30700. .30894
/rpt_family="L2"
repeat_region      30992. .31130
/rpt_family="MER45"
repeat_region      complement(31153. .31285)
/rpt_family="MIR"
```

Query Match 2.8%; Score 54.2; DB 86; Length 173126;
Best Local Similarity 49.1%; Pred. No. 0.012;
Matches 143; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

```
QY 30 GGACAGTGTCCGAGTCCCGCGGTGCGCCAGAGCGCTCGAGAGCCCGAGCCCGGTG 89
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163150 GGCGCCGAGGGGGGAGGGCGAGGAGGAGGAGGAGCCCGCGCGCGCGAGCCCGC 163209
QY 90 CCGGTGCTCTCGCGCCCGCGAGTCTCCGCGTACGCTCATTCACAGACACGCTGTTGGA 149
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163210 CCGCGCGCGCGCGAGCGCGGACGCTGTGCGCGCGCTCGAGGTTGAGCCCGAGCGTCC 163269
QY 150 CTTGTGTCAGTGACAGCGACGAGAGGTCTTGGAAATCGACACCCAGTACAGTGCCTG 209
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163270 GGAAGCGAGGACGAGCGCGCGCGCGCGGCGTGCAGAGCGCGCGCGCGCGCGCG 163329
QY 210 CGCGCGCGCGCGCGCGCGCGCTAAACCTGAGCAGGAGCAGCAGTACAGTGAAGGGGC 269
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163330 CCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 163389
QY 270 GGCGGAGGGGCGTCCGCGGAGCGCGCGTACATTTGTCGACGCGCGCGCGCGCGCG 320
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163390 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 163440
```

```
RESULT 9
AC074153 AC074153 224583 bp DNA 15-JUL-2000
LOCUS Mus musculus clone RP23-172E18, WORKING DRAFT SEQUENCE, 39
DEFINITION unordered pieces.
ACCESSION AC074153 GI:9211263
VERSION AC074153.1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 224583)
REFERENCE DOE Joint Genome Institute.
AUTHORS Sequencing of Mouse
TITLE Unpublished
JOURNAL 2 (bases 1 to 224583)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-23_172E18
-----
Summary Statistics
Consensus quality: 142282 bases at least Q40
Consensus quality: 169507 bases at least Q30
Consensus quality: 182683 bases at least Q20
Estimated insert size: 207300; agarose-fp estimation
Estimated insert size: 220783; sum-of-contigs estimation
Quality coverage: 3.11 in Q20 bases; agarose-fp estimation
```


AUTHORS
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Deatellano, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Garlynd, J., Grant, G., Hago, B., Heaford, A., Horton, L., Howard, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehocsky, J., Levine, R., Lieu, G., Locke, K., Macdonald, P., Marcus, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, R. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Turrell, A., Vassiliou, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zody, M.

TITLE
 Direct Submission

JOURNAL
 Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 9, 2000 this sequence version replaced gi:7331400.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L3197
 Center clone name: 193_P_5
 ----- Summary Statistics
 Sequencing vector: M13: M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 202781 bases at least Q40
 Consensus quality: 209320 bases at least Q30
 Consensus quality: 211305 bases at least Q20
 Insert size: 212000; agarose-fp
 Insert size: 212274; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; agarose-fp
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1
 * 1375 1474: contig of 1374 bp in length
 * 1375 1474: gap of 100 bp
 * 1475 2242: contig of 768 bp in length
 * 2243 2342: gap of 100 bp
 * 2343 5201: contig of 2859 bp in length
 * 5202 5301: gap of 100 bp
 * 5302 7336: contig of 2035 bp in length
 * 7337 7436: gap of 100 bp
 * 7437 10614: contig of 3178 bp in length
 * 10615 10714: gap of 100 bp
 * 10715 14387: contig of 3673 bp in length
 * 14388 14487: gap of 100 bp
 * 14488 18778: contig of 4291 bp in length
 * 18779 18878: gap of 100 bp
 * 18879 24284: contig of 5406 bp in length
 * 24285 24384: gap of 100 bp
 * 24385 29902: contig of 5518 bp in length
 * 29903 30002: gap of 100 bp
 * 30003 36354: contig of 6352 bp in length
 * 36355 36454: gap of 100 bp
 * 36455 43626: contig of 7172 bp in length

FEATURES
 source
 * 43627 43726: gap of 100 bp
 * 43727 50099: contig of 6373 bp in length
 * 50100 50199: gap of 100 bp
 * 50200 56326: contig of 6127 bp in length
 * 56327 56426: gap of 100 bp
 * 56427 64338: contig of 7912 bp in length
 * 64339 64438: gap of 100 bp
 * 64439 72117: contig of 7679 bp in length
 * 72118 72217: gap of 100 bp
 * 72218 82743: contig of 10566 bp in length
 * 82744 82843: gap of 100 bp
 * 82844 93078: contig of 10235 bp in length
 * 93079 93178: gap of 100 bp
 * 93179 103611: contig of 10433 bp in length
 * 103612 103711: gap of 100 bp
 * 103712 114971: contig of 11260 bp in length
 * 114972 115071: gap of 100 bp
 * 115072 128960: contig of 13909 bp in length
 * 128961 129080: gap of 100 bp
 * 129081 142071: contig of 12991 bp in length
 * 142072 142171: gap of 100 bp
 * 142172 165935: contig of 23764 bp in length
 * 165936 166035: gap of 100 bp
 * 166036 190737: contig of 24702 bp in length
 * 190738 190837: gap of 100 bp
 * 190838 214574: contig of 23737 bp in length.
 Location/Qualifiers
 1. 214574
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /map="11"
 /clone.lib="RP23-193P5"
 /clone.lib="RPCT-23 Female Mouse BAC"
 1. 1374
 /note="assembly-fragment"
 1475. 2242
 /note="assembly-fragment"
 clone_end:SP6
 vector_side:left
 2343. 5201
 /note="assembly-fragment"
 5302. 7336
 /note="assembly-fragment"
 7437. 10614
 /note="assembly-fragment"
 10715. 14387
 /note="assembly-fragment"
 14488. 18778
 /note="assembly-fragment"
 18879. 24284
 /note="assembly-fragment"
 24385. 29902
 /note="assembly-fragment"
 30003. 36354
 /note="assembly-fragment"
 36455. 43626
 /note="assembly-fragment"
 43727. 50099
 /note="assembly-fragment"
 50200. 56326
 /note="assembly-fragment"
 56427. 64338
 /note="assembly-fragment"
 64439. 72117
 /note="assembly-fragment"
 72218. 82743
 /note="assembly-fragment"
 82844. 93078
 /note="assembly-fragment"
 93179. 103611
 /note="assembly-fragment"
 103712. 114971
 /note="assembly-fragment"

	misc_feature	/note="assembly_fragment" 115072..118960 /note="assembly_fragment"
	misc_feature	/note="assembly_fragment" 129081..142071 /note="assembly_fragment"
	misc_feature	142172..165935 /note="assembly_fragment"
	misc_feature	166036..180737 /note="assembly_fragment"
	misc_feature	190838..214574 /note="assembly_fragment"
		clone_end::T7
		vector_side:right"
BASE COUNT	56656 a 49877 c 51317 g 54418 t 2306 others	
ORIGIN		
Query Match	2.5% Score 49.2; DB 65; Length 214574;	
Best Local Similarity	50.4%; Pred No. 0.18;	
Matches 120; Conservative	0; Mismatches 118; Indels 0; Gaps 0	
OY	147 GGACTTGTCAGTCGACAGCGACGAAGAGCTCTTGAAATCGCAGACCAGTAGAGGTGCC	206
Db	189074 GGGCAGGAGCAGAGAAGCGAGGCAGAGCCAGAGCGCAGAGGCGCAGAGGCGCACAGGC	189015
OY	207 GTTGCCCCGCTCCCGCCGCGCTTAACCTTAGCAGCAGCAGCGACGTGACAGTAAAG	266
Db	189014 AGGGGCAGAGCGCAGAGGCGAGGGCGAAGAGCAGGGCGCAGAGGCGCAGAGAGGC	188955
OY	267 GGGCGCCGAGGGGCGCTGGCGAGACCCCGGCTCATTTGTCGCAACGCGCGCGCGCGCT	326
Db	188954 AGGGCAGAGAGGCGCAGGGCAGGGCGCAGGGCGCAGGGCGCAGAGCGCAGAGCGCAGGG	188895
OY	327 GCTGATCCCGAGAGCGCGCGGTGGTCCAGTGTACTCTCCGGAAGGTACAGACGAC	384
Db	188894 GCAGAAAGCGCAGAGAGCGGGGCGAGGCGCAGAGGCGCAGAGGCGCAGAGAGCGC	188837
RESULT 12		
LOCUS	SCPSCL	11696 bp DNA BCT 30-JUN-1993
DEFINITION	S.clavuligerus linear plasmid pSCL (complete sequence).	
ACCESSION	X54107 S51598	
VERSION	X54107.1 GI:48758	
KEYWORDS	Inverted repeat; linear plasmid.	
SOURCE	Streptomyces clavuligerus.	
ORGANISM	Plasmid Streptomyces clavuligerus	
	Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:	
	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.	
REFERENCE	1 (bases 1 to 11696)	
AUTHORS	Roy,K.L.	
TITLE	Direct Submission	
JOURNAL	Submitted (26-JUN-1990) Roy K.L., Dept of Microbiology, University of Alberta, Biological Sciences Bldg Rm M330, Edmonton Alberta, Canada T6G 2E9	
REFERENCE	2 (bases 1 to 11696)	
AUTHORS	Wu,X. and Roy,K.L.	
TITLE	Complete nucleotide sequence of a linear plasmid from Streptomyces clavuligerus and characterization of its RNA transcripts	
JOURNAL	J. Bacteriol. 175 (1), 37-52 (1993)	
MEDLINE	93106972	
FEATURES	Location/Qualifiers	
source	1..11696	
	/organism="Streptomyces clavuligerus"	
	/plasmid="pSCL"	
	/strain="NRRL 3585"	
	/db_xref="taxon:1901"	
	complement(1446..2237)	
	/note="ORF1"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="CAA38040.1"	
	/db_xref="GI:48759"	

CDS

/db_xref="SPTREMBL:O05073"
/translation="MAATRDALINQKGVACTTTTHLGITGLAAAGRTLLVDLDGOG
/transl_table=1
/protein_id="CAA38041.1"
/db_xref="GI:581632"
/db_xref="SPTREMBL:O05074"
/translation="MSMSEELRAEVEAEADVIGTLRPEARDIVDLYSALAPTDT
FREPRIGELMDLILAMRGTPCDPVITYSAVLRKGDLERIGGAWMLHALDVPTTR
NALPAEVLAEESRRALLAAGVTAVTAQOREGTIAAALPATAVGAAREVNDRLTAA
DAPDAIDLFDQADDEPEMVPGVTAFMDRLMITACGGGKSLLRLIIVRAAGLGG
PMRRRIEYEVRLVLAENAAASOVPRMYLRMAAAAAGAFAVGARMHIEVRAGGLD
TTARAWMTARTVERVRPDIMSIGPLYLTSSGADEEARPLMALEMYQAISDGAA
MLELHSHPAAGVRRDRIPGSSLRMWEFEGGSLSPATGMCAEBELRLMDMLPWG
PSEKEMWPQCCEGGTWPMTAYAGNVPITPATLTTERQAAAGLIPROPFWDDQQ
G"

CDS

complement(6349..6786)
/note="ORF5"
/codon_start=1
/transl_table=1
/protein_id="CAA38042.1"
/db_xref="GI:48761"
/db_xref="SPTREMBL:O05076"
/translation="MEALLHLRFALROHGEMFEDSGDNPLAKYEAITTAGLPETPAQ
OSHHVIAPRPVSVDARGCGHELSFHFRGATGCYACVADMELMOCSGRFPAAEDWMH
VSMPCARICRHSAQDHGLGEBCSCSKGACCITHPSASGVAR"

CDS

complement(6931..7914)
/note="large RNA transcript"
complement(7139..7885)
/note="ORFL"
/codon_start=1
/transl_table=1
/protein_id="CAA38043.1"
/db_xref="GI:48762"
/db_xref="SPTREMBL:O05077"
/translation="MPKITERPLPPYMOVIAHTRIDIDSGELRACGLIPSDNNIAQEMOE
ISRVAOKYTTLKAOGILVEAVOGSGTYRAAHDHLHRSGDRAASVRRGTITAGC
YARIYSABELARPADYAEVLSQDECAATIRLRITYNENOPIASSTSWYGNLAAAA
PYLULKTERKEGSWAYLEBOTGMTAVTCQDMRTSARMATEQDAQULGITELPAIKTEART
ILRDANSVAEEVINGINSREGSIYDYVS"

promoter

complement(7921..7951)
/note="promoter for large RNA transcript"

prim_transcript

complement(7951..8400)
/note="small RNA transcript"
8114..9367
/note="ORF3"
/codon_start=1
/transl_table=1
/protein_id="CAA38044.1"
/db_xref="GI:581633"
/db_xref="SPTREMBL:O05075"
/translation="MLEAPRLPSDPTCLARSQSVIAERRVAVQGVPRPVAGLAEE
VEMEREEDAQPTDPGGGVGRMSBERBELGARAVAPVGOGRTPSYTRAARDVLDL
TAEVPLTTPYLARAALVEEHGDHLVRYVARLRDLDAEWATHAREDAVOYMLDYVAA
GVVELLEVPDLWPRLAAAAKWSLLDNTFTRRRREWLVRPEDRSDEVEALACG
GDSITYCADELMEPEPEGGWACVAYERIALAPRQREVELCTEGMTTPAIAAA
LCISRSVDLARFAVTALGSPGYVRORSRCAGCPDPAGERVYLDRLPNKTORDVYA
LRAGGSFELEGOLGHRGYAHELHYTALSRLREMWDQRDLDPVPAPARSKPKPHH
APAGCARNSCVSGCYLTRATGCAAS"

CDS

complement(8411..8442)
/note="promoter for small RNA transcript"
9696..10679
/note="ORF4"
/codon_start=1
/transl_table=1
/protein_id="CAA38045.1"
/db_xref="GI:48764"

Promoter

CDS

```

/db_xref="SPTREMBL:005083"
/translation="MGRRETEPCVPRPQWVLRLLIGGAVAVAGVAMAASVTLAELGHA
VGVWMDGRSLNSLPVAVDLALVAGVAVGSPMTARSRLAGITITGAAGSVLNA
IGHLVESGDKKVPILRIAVASVAPPLVAVLGVVLAIPVAGAGCOCOAELERD
REARITSAHMDPADGTADRHDSITARNAPAPPRAPADAVYRPAAPAAARTRAA
DDPAPVPAAPAAOTGPVRYTARLGDPRDSDQIWERPAPVDRDPLDPCADPTPEP
LLTTPDSRRARILMALONTDMTKRIGEFAGRSASVYNTKKAIRINDGILI"
BASE COUNT      1590 a      4235 c      4179 g      1692 t
ORIGIN

Query Match      2.5%; Score 48.2; DB 3; Length 11696;
Best Local Similarity 55.8%; Pred. No. 0.41;
Matches 92; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY 196 GTACAGTGCCTGCTGCGCCCTCCCGCGCGCTAACCTGACGAGCAGCAGCAGT 255
    || || || || || || || || || || || || || || || || || || || ||
DB 8198 GTGGCGCAGGTGGGTGTCGGCCGCGTGTGCTGGCTGGCTGAGCAGTGCAGCAGTGG 8257
    || || || || || || || || || || || || || || || || || || || ||
OY 256 GACAGTGAAGGGCGCGCGAGGCGCTCGCGAGCCCGCGCTACATTGTGCGAGCGCGG 315
    || || || || || || || || || || || || || || || || || || || ||
DB 8258 GAGCGGGAGCGAGCAGCGAGCGCCGAGCGAGCGCGGAGGTGGGTGCTGTCGCGTGG 8317
    || || || || || || || || || || || || || || || || || || || ||
OY 316 CGCGCGCGCGCTGTCGATCCCGAGAGCGCGCGTGTCCAGTGG 360
    || || || || || || || || || || || || || || || || || || || ||
DB 8318 TCGGAGCGCTCAGCGAGCTGGGCGCGCGCGCGTGTGCGCGTGG 8362
    || || || || || || || || || || || || || || || || || || || ||

RESULT 13
AC069456      160893 bp      DNA      HTG      25-JUN-2000
LOCUS      Homo sapiens chromosome 4 clone RP11-280K20 map 4, WORKING DRAFT
DEFINITION      AC069456
SEQUENCE      22 unordered pieces.
AC069456
AC069456.2 GI:8705194
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
1 (bases 1 to 160893)
JOURNAL      Homo sapiens chromosome 4, clone RP11-280K20
REFERENCE      Unpublished
AUTHORS      2 (bases 1 to 160893)
      Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
      Anderson, S., Baldwin, J., Barro, N., Bastien, V., Beda, F.,
      Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
      Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
      Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, D. S.,
      Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
      Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
      Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
      Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kahn, L., Karatas, A.,
      Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehocsky, J.,
      Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
      McCathly, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
      Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,
      Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
      O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
      Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
      Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
      Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
      Tassif, S., Theodore, J., Turrell, A., Travers, M., Trigilio, J.,
      Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
      Young, G., Zainoun, J., Zimmer, A. and Zody, M.
TITLE      Direct Submission
JOURNAL      Submitted (30-MAY-2000) Whitehead Institute/MIT Center for Genome
COMMENT      Research, 330 Charles Street, Cambridge, MA 02141, USA
      On Jan 25, 2000 this sequence version replaced gi:8103949.
      All repeats were identified using RepeatMasker:
      Smit, A.F.A. & Green, P. (1996-1997)
      http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: RIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9896
Center clone name: 280_K-20
----- Summary Statistics
Sequencing vector: M13; M7815: 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 152447 bases at least Q40
Consensus quality: 156415 bases at least Q30
Consensus quality: 157976 bases at least Q20
Insert size: 163000; agarose-fp
Insert coverage: 158793; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1360: contig of 1360 bp in length
1361 1460: gap of 100 bp
1461 2527: contig of 1067 bp in length
2528 2627: gap of 100 bp
2628 3976: contig of 1349 bp in length
3977 4076: gap of 100 bp
4077 6740: contig of 2664 bp in length
6741 6840: gap of 100 bp
6841 9482: contig of 2642 bp in length
9483 9582: gap of 100 bp
9583 12171: contig of 2589 bp in length
12172 12271: gap of 100 bp
12272 15766: contig of 3495 bp in length.
15767 15866: gap of 100 bp
15867 19727: contig of 3861 bp in length
19728 19827: gap of 100 bp
19828 24165: contig of 4338 bp in length
24166 24265: gap of 100 bp
24266 28290: contig of 4025 bp in length
28291 28390: gap of 100 bp
28391 33226: contig of 4836 bp in length
33227 33326: gap of 100 bp
33327 40113: contig of 6787 bp in length
40114 40213: gap of 100 bp
40214 45498: contig of 5285 bp in length
45499 45598: gap of 100 bp
45599 52958: contig of 7360 bp in length
52959 53058: gap of 100 bp
53059 61457: contig of 8399 bp in length
61458 61557: gap of 100 bp
61558 70503: contig of 8946 bp in length
70504 70603: gap of 100 bp
70604 79766: contig of 9163 bp in length
79767 79866: gap of 100 bp
79867 92993: contig of 13127 bp in length
92994 93093: gap of 100 bp
93094 108057: contig of 14964 bp in length
108058 108157: gap of 100 bp
108158 126261: contig of 18104 bp in length
126262 126361: gap of 100 bp
126362 142680: contig of 16319 bp in length
142681 142780: gap of 100 bp
142781 160893: contig of 18113 bp in length.
Location/Qualifiers
1. 160893
FEATURES
source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-280K20"
/clone_lib="RPC1-11 Human Male BAC"
1. .1360
misc_feature
/note="assembly-fragment"
1461. .2527
misc_feature
/note="assembly-fragment"
2628. .3976
misc_feature
/note="assembly-fragment
clone_end:17
vector_side:right"
4077. .6740
misc_feature
/note="assembly-fragment"
6841. .9482
misc_feature
/note="assembly-fragment"
9583. .12171
misc_feature
/note="assembly-fragment"
12272. .15766
misc_feature
/note="assembly-fragment"
15867. .19727
misc_feature
/note="assembly-fragment"
19828. .24165
misc_feature
/note="assembly-fragment"
24266. .28290
misc_feature
/note="assembly-fragment"
28391. .33326
misc_feature
/note="assembly-fragment"
33327. .40113
misc_feature
/note="assembly-fragment"
40214. .45498
misc_feature
/note="assembly-fragment"
45599. .52958
misc_feature
/note="assembly-fragment"
53059. .61457
misc_feature
/note="assembly-fragment"
61558. .70503
misc_feature
/note="assembly-fragment"
70604. .79766
misc_feature
/note="assembly-fragment"
79867. .92993
misc_feature
/note="assembly-fragment"
93094. .108057
misc_feature
/note="assembly-fragment"
108158. .126261
misc_feature
/note="assembly-fragment"
126362. .142680
misc_feature
/note="assembly-fragment
clone_end:5p6
vector_side:left"
142781. .160893
misc_feature
/note="assembly-fragment"
BASE COUNT 51136 a 28485 c 27792 g 51377 t 2103 others
ORIGIN

Query Match 2.4% Score 47.4; DB 74; Length 160893;
Best Local Similarity 46.3%; Pred. No. 0.5;
Matches 156; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

Dy 103 GCCGGGAGTCTCCGGCTAGCTCAATTCACAGACCGTCTTGAGACTTGTCAGTAC 162
|||||
Db 91957 GCTTCACATCTTCGGCCTTGGCGCGGACCTCTCAAGCCCTTGGGTGGATGGCACTG 92016

Dy 163 AGCGACGAGAGGTCTTGAGAACTCCAGACCCAGTAGAGGTCCGAGTCCCGCCCTCCC 222
|||||
Db 92017 AGTGCCTGGAGACACAGGGGCGCGCTCGTTGGGGGGCGCTCTTGGGCGCGCTCTT 92076

Dy 223 GCGCGGCTAAACCTTGAGACGACGACGACAGTACAGTGAAGGGGCGCGGAGGGCCT 282
|||||
Db 92077 TGGGAGGCTCCCGCGCACGATCCCAACGAGGAGGGGTGAGGCTTCAGGCATGGCGGCT 92136

```

QY	283	GGGGAGCCCGCCGATCAATTGGTGGAGCGGCGCGCGCTGCTGGATCCCGAGAG	342
Db	92137	GGAGTCCGGAGACCTCTGCCCTCGCGGGAGAGCGATTAAGCCCGGTGAATAATCCAGCCCA	92196
QY	343	GGCGCGGTGTCGCCAGTGTATACCTCGGAGAGACAGACAGCCTTCACCTATTCACAT	402
Db	92197	GGCGTGTGTGGCGGCGACTGCTGTGGGGAGACACAGACACCTCTCCGACGCGGTGGCGCGC	92256
QY	403	AATTCATCCCTCTTGAACACTGTGCCCTTCAGACCTG	439
Db	92257	TGCTAAGCCCTCATTTGCTGTGGCGCTTTAGGCGCG	92293
RESULT	14		
LOCUS	AC022573/c		
DEFINITION	Homo sapiens clone RP11-12N11, LOW-PASS SEQUENCE SAMPLING.		
ACCESSION	AC022573		
VERSION	AC022573.2		
KEYWORDS	HTG; HTGS_PHASE0.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A., Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearbell, K., Dewar, K., Domino, M., Doyle, M., Fenesator, J., Ferrelia, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, D., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kan, L., Kartas, A., Klein, J., Landers, T., Lehozky, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McCurk, A., McKernan, K., McPheters, R., Meldrim, J., Menus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zody, M.		
TITLE	2 (bases 1 to 77836)		
JOURNAL	Unpublished		
REFERENCE	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A., Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearbell, K., Dewar, K., Domino, M., Doyle, M., Fenesator, J., Ferrelia, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, D., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kan, L., Kartas, A., Klein, J., Landers, T., Lehozky, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McCurk, A., McKernan, K., McPheters, R., Meldrim, J., Menus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zody, M.		
COMMENT	Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Mar 3, 2000 this sequence version replaced gi:6910780. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html Center: Whitehead Institute/ MIT Center for Genome Research Genome Center Center code: WIMR Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu Project Information Center project name: L3299 Center clone name: 12_N_11 NOTE: This record contains 89 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will		

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web Site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: I3299
Center clone name: 12_N_11

* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

- * will be sequenced to completion. In the event that
- * the record is updated, the accession number will

* be preserved.
* 1 762: contig of 762 bp in length
* 763 862: gap of 100 bp
* 863 1633: contig of 771 bp in length
* 1634 1733: gap of 100 bp
* 1734 2505: contig of 772 bp in length
* 2506 2605: gap of 100 bp
* 2606 3373: contig of 768 bp in length
* 3374 3473: gap of 100 bp
* 3474 4273: contig of 800 bp in length
* 4274 4373: gap of 100 bp
* 4374 5153: contig of 780 bp in length
* 5154 5253: gap of 100 bp
* 5254 6016: contig of 763 bp in length
* 6017 6116: gap of 100 bp
* 6117 6881: contig of 765 bp in length
* 6882 6981: gap of 100 bp
* 6982 7736: contig of 757 bp in length
* 7737 7838: gap of 100 bp
* 7839 8600: contig of 762 bp in length
* 8601 8700: gap of 100 bp
* 8701 9464: contig of 764 bp in length
* 9465 9564: gap of 100 bp
* 9565 10338: contig of 775 bp in length
* 10340 10439: gap of 100 bp
* 10440 11223: contig of 784 bp in length
* 11224 11323: gap of 100 bp
* 11324 12111: contig of 788 bp in length
* 12112 12211: gap of 100 bp
* 12212 12970: contig of 759 bp in length
* 12971 13070: gap of 100 bp
* 13071 13870: contig of 800 bp in length
* 13871 13970: gap of 100 bp
* 13971 14724: contig of 754 bp in length
* 14725 14824: gap of 100 bp
* 14825 15596: contig of 772 bp in length
* 15597 15696: gap of 100 bp
* 15697 16456: contig of 760 bp in length
* 16457 16556: gap of 100 bp
* 16557 17314: contig of 758 bp in length
* 17315 17414: gap of 100 bp
* 17415 18176: contig of 762 bp in length
* 18177 18276: gap of 100 bp
* 18277 19036: contig of 760 bp in length
* 19037 19136: gap of 100 bp
* 19137 19931: contig of 795 bp in length
* 19932 20031: gap of 100 bp
* 20032 20823: contig of 792 bp in length
* 20824 20923: gap of 100 bp
* 20924 21708: contig of 785 bp in length
* 21709 21808: gap of 100 bp
* 21809 22582: contig of 774 bp in length
* 22583 22682: gap of 100 bp
* 22683 23443: contig of 761 bp in length
* 23444 23543: gap of 100 bp
* 23544 24326: contig of 783 bp in length
* 24327 24426: gap of 100 bp
* 24427 25200: contig of 774 bp in length
* 25201 25300: gap of 100 bp
* 25301 26065: contig of 765 bp in length
* 26066 26165: gap of 100 bp
* 26166 26922: contig of 757 bp in length
* 26923 27022: gap of 100 bp
* 27023 27796: contig of 774 bp in length
* 27797 27896: gap of 100 bp
* 27897 28680: contig of 784 bp in length
* 28681 28780: gap of 100 bp
* 28781 29571: contig of 791 bp in length
* 29572 29671: gap of 100 bp
* 29672 30473: contig of 802 bp in length
* 30474 30573: gap of 100 bp
* 30574 31338: contig of 765 bp in length
* 31339 31438: gap of 100 bp

* 31439 32217: contig of 779 bp in length
* 32218 32317: gap of 100 bp
* 32318 33093: contig of 776 bp in length
* 33094 33193: gap of 100 bp
* 33194 33956: contig of 763 bp in length
* 33957 34056: gap of 100 bp
* 34057 34818: contig of 762 bp in length
* 34819 34918: gap of 100 bp
* 34919 35811: contig of 893 bp in length
* 35812 35911: gap of 100 bp
* 35912 36663: contig of 752 bp in length
* 36664 36763: gap of 100 bp
* 36764 37553: contig of 790 bp in length
* 37554 37653: gap of 100 bp
* 37654 38412: contig of 759 bp in length
* 38413 38512: gap of 100 bp
* 38513 39258: contig of 747 bp in length
* 39260 39359: gap of 100 bp
* 39360 40144: contig of 785 bp in length
* 40145 40244: gap of 100 bp
* 40245 41027: contig of 783 bp in length
* 41028 41127: gap of 100 bp
* 41128 41907: contig of 780 bp in length
* 41908 42007: gap of 100 bp
* 42008 42777: contig of 770 bp in length
* 42778 42877: gap of 100 bp
* 42878 43682: contig of 805 bp in length
* 43683 43782: gap of 100 bp
* 43783 44559: contig of 777 bp in length
* 44560 44659: gap of 100 bp
* 44660 45429: contig of 770 bp in length
* 45430 45529: gap of 100 bp
* 45530 46280: contig of 751 bp in length
* 46281 46380: gap of 100 bp
* 46381 47146: contig of 766 bp in length
* 47147 47246: gap of 100 bp
* 47247 48008: contig of 762 bp in length
* 48009 48108: gap of 100 bp
* 48109 48881: contig of 773 bp in length
* 48882 48981: gap of 100 bp
* 48982 49762: contig of 781 bp in length
* 49763 49862: gap of 100 bp
* 49863 50654: contig of 792 bp in length
* 50655 50754: gap of 100 bp
* 50755 51535: contig of 781 bp in length
* 51536 51635: gap of 100 bp
* 51636 52451: contig of 816 bp in length
* 52452 52551: gap of 100 bp
* 52552 53307: contig of 756 bp in length
* 53308 53407: gap of 100 bp
* 53408 54173: contig of 766 bp in length
* 54174 54273: gap of 100 bp
* 54274 55037: contig of 764 bp in length
* 55038 55137: gap of 100 bp
* 55138 55904: contig of 767 bp in length
* 55905 56004: gap of 100 bp
* 56005 56753: contig of 749 bp in length
* 56754 56853: gap of 100 bp
* 56854 57637: contig of 784 bp in length
* 57638 57737: gap of 100 bp
* 57738 58519: contig of 782 bp in length
* 58520 58619: gap of 100 bp
* 58620 59414: contig of 795 bp in length
* 59415 59514: gap of 100 bp
* 59515 60301: contig of 787 bp in length
* 60302 60401: gap of 100 bp
* 60402 61193: contig of 792 bp in length
* 61194 61293: gap of 100 bp
* 61294 62068: contig of 775 bp in length
* 62069 62168: gap of 100 bp

Query Match 2.48; Score 47.2; DB 67; Length 77836;
Best Local Similarity 43.0%; Pred. No. 0.6;

	Matches	148:	Conservative	0:	Mismatches	196:	Indels	0:	Gaps	0:
QY	30	GGGACGTGgTCCGAGATCCCGCGGTgGCCCGGACGCTCGAGAGACCCCGAGCGCCCTGTG	89							
Db	35624	GGGCGGGGGGGGGCGGNNNNNNGGGGGGGGCGCGCGGCGCGGCGCGGGCGGGGCGGGG	35655							
QY	90	CCGGTGTCTCTCGCGCCCGGCAATCTCCGGGCTTAgGCTCAATTCCAGACACCGGTCTTGTGA	149							
Db	35564	CCGGCGCCCGCGCGGG	35505							
QY	150	CTTGTGCTgCATGTCACAGCAGAGAGGTCTTGGAAgTCGCGAGACCCAGTAgGATGCGCGGT	209							
Db	35504	GGGGGGGGGGGGCGCGCGGNNGGGAGGGGGGGCGCGGGGGGGGGGGGGGGGGGGGGGGG	35445							
QY	210	CGCCCGCTCCCGCGCGCGGCTAAACTTgAGCAGGACACGCACTGACAgTGAAGGGGGC	269							
Db	35444	CCCGGGGGGGCGGGGGGGGGGGCGGGGCGGNNGGNNGGNNGGGGGGGGGGGGGGGGGGGGG	35385							
QY	270	GAGCGAGAGGGCTGCGGGAGACCCGCGTACATTTgGTGCGACGGCGGGCGGGCGGTCT	329							
Db	35384	GGGGGGGGGGGNNNGGGGGGGGGGGGGCGCGGGNNNGGGGGGGGGGGGCGGGGGGGGG	35325							
QY	330	GGATCCCGAGAGGGCGCGGTGTCCCAgTACTCCGGGAAG	373							
Db	35324	GGGCGCGGGGNCGGGGGGGGGGCGNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	35281							

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vetta, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Wellington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.
and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 147225)
Morley, K.C.
Direct Submission
Submitted (05-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 3, 2000 this sequence version replaced g1:9438272.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HADJ
Center clone name: RP11-430L6
----- Summary Statistics
Sequencing vector: M13, L08821

RESULT 15	LOCUS	AC022498	DEFINITION	ACCESSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
15	LOCUS	AC022498	147225 bp DNA	HTG	04-NOV-2000				
	LOCUS	AC022498	147225 bp DNA	HTG	04-NOV-2000				
	DEFINITION	AC022498	147225 bp DNA	HTG	04-NOV-2000				
	ACCESSION	AC022498	147225 bp DNA	HTG	04-NOV-2000				
	KEYWORDS	AC022498	147225 bp DNA	HTG	04-NOV-2000				
	SOURCE	AC022498	147225 bp DNA	HTG	04-NOV-2000				
	ORGANISM	AC022498	147225 bp DNA	HTG	04-NOV-2000				
	REFERENCE	AC022498	147225 bp DNA	HTG	04-NOV-2000				
	AUTHORS	AC022498	147225 bp DNA	HTG	04-NOV-2000				

```

Assembly program: Phrap, version 0.990329
Consensus quality: 119390 bases at least Q40
Consensus quality: 134432 bases at least Q30
Consensus quality: 138749 bases at least Q20
Estimated insert size: 138179; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
      (see http://www.bpsc.bcm.tmc.edu/docs/genbank\_difft\_data.html)
NOTE: This is a 'working draft' sequence. It currently
      consists of 29 contigs. The true order of the pieces
      is not known and their order in this sequence record is
      arbitrary. Gaps between the contigs are represented as
      runs of N, but the exact sizes of the gaps are unknown.
      This record will be updated with the finished sequence
      as soon as it is available and the accession number will
      be preserved.

1
10679      10678: contig of 10678 bp in length
10679      10778: gap of unknown length
10779      27896: contig of 17118 bp in length
17897      27996: gap of unknown length
27997      40116: contig of 12120 bp in length
40117      40216: gap of unknown length
40217      53842: contig of 13626 bp in length
53843      53942: gap of unknown length
53943      60133: contig of 6191 bp in length
60134      60233: gap of unknown length
60234      66953: contig of 6720 bp in length
66954      76053: gap of unknown length
76055      72921: contig of 5868 bp in length
72922      73021: gap of unknown length
73022      81158: contig of 8137 bp in length
81159      81258: gap of unknown length
81259      88668: contig of 7410 bp in length
88669      88668: gap of unknown length
88769      94722: contig of 5954 bp in length
94723      94822: gap of unknown length
94823      100284: contig of 5462 bp in length
100285      100384: gap of unknown length
100385      105379: contig of 4995 bp in length
105380      105479: gap of unknown length
105480      110304: contig of 4825 bp in length
110305      110404: gap of unknown length
110405      116806: contig of 6402 bp in length
116807      116906: gap of unknown length

```



```
* 116907 120759: contig of 3853 bp in length
* 120760 120859: gap of unknown length
* 120860 125348: contig of 4489 bp in length
* 125349 125448: gap of unknown length
* 125449 127226: contig of 1778 bp in length
* 127227 127326: gap of unknown length
* 127327 129574: contig of 2248 bp in length
* 129575 129674: gap of unknown length
* 129675 131343: contig of 1669 bp in length
* 131344 131443: gap of unknown length
* 131444 133042: contig of 1599 bp in length
* 133043 133142: gap of unknown length
* 133143 135153: contig of 2011 bp in length
* 135154 135253: gap of unknown length
* 135254 137006: contig of 1753 bp in length
* 137007 137106: gap of unknown length
* 137107 138550: contig of 1444 bp in length
* 138551 138650: gap of unknown length
* 138651 140149: contig of 1499 bp in length
* 140150 140249: gap of unknown length
* 140250 141631: contig of 1382 bp in length
* 141632 141731: gap of unknown length
* 141732 143173: contig of 1442 bp in length
* 143174 143273: gap of unknown length
* 143274 144588: contig of 1315 bp in length
* 144589 144688: gap of unknown length
* 144689 145782: contig of 1094 bp in length
* 145783 145882: gap of unknown length
* 145883 147225: contig of 1343 bp in length.
```

FEATURES

Source

```
1. 147225
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="3"
   /clone="RP11-430L16"
```

BASE COUNT 41313 a 31018 c 30751 g 41299 t 2844 others

ORIGIN

Query Match 2.4%; Score 47; DB 67; Length 147225;

Best Local Similarity 49.3%; Pred. No. 0.63; Mismatches 151; Indels 1; Gaps 1;

Matches 148; Conservative 0; Mismatches 151; Indels 1; Gaps 1;

```
QY 25 CTGAGGGACGTGTCGAGGTCCGCGGTGCGAGCGCTCGAGAGCCCGGAGCCGCC 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88345 CGGCGGGGGGGGGGGGGGGGGTGGGGTGGCGGGGGCCGGGGGGGGCGCGCG 88404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 85 CGTGGCGGTCTCTCCGCCGCCGCACTCTCCGCTAGGCTCATTCAGACACCGTCTT 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88405 CGGCGGGGGGGGGGGGGGGGCTTTCGGGGGCCCGCGGGGGGGGGCGGT 88464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 145 GTGACTTGTCAGTGCAGCAGAGAGAGTCTTGGAGTCGAGACCCAGTAGAGTG 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88465 GCCCGGTGGGCCCGGGGGGGGGGGGGGGGGTGGGGGGGGCGTCCCGCGCGCG 88524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 205 CCGGTCGCCCGCTCCCGCGCGCGCTAAACTGACAGCAGACAGTACAGTAA 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88525 GCGGGGGGGCGG -CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGGTG 88583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 265 GGGCGGGCGCGAGGGGCTCGGGGAGCCCGGTACATTGGTGCAGCGGGCGGGCGG 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88584 GCGGTGGCGGNGGCGGGCGGTGGGGTCCGCGCGCGCGCGCGCGCGCGCGTGGCGG 88643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: September 11, 2001, 12:44:19
Job time: 3951 sec

This Page Blank

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2001, 11:41:48 ; Search time 1471.75 Seconds
(without alignments)
12498.898 Million cell updates/sec

Title: US-09-192-611-1

Perfect score: 1946
Sequence: 1 ACAGTGTGGACATGCGGA.....CGCGATCGCGCCGACTCGAG 1946

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*
48: gb_est48.*
49: gb_est49.*
50: gb_est50.*
51: gb_est51.*
52: gb_est52.*
53: gb_est53.*
54: gb_est54.*
55: gb_est55.*
56: gb_est56.*
57: gb_est57.*
58: gb_est58.*
59: gb_est59.*
60: gb_est60.*
61: gb_est61.*
62: gb_est62.*
63: gb_est63.*
64: gb_est64.*
65: gb_est65.*
66: gb_est66.*
67: gb_est67.*
68: gb_est68.*
69: gb_est69.*
70: gb_est70.*
71: gb_est71.*
72: gb_est72.*
73: gb_est73.*
74: gb_est74.*
75: gb_est75.*
76: gb_est76.*
77: gb_est77.*
78: gb_est78.*
79: gb_est79.*
80: gb_est80.*
81: gb_est81.*
82: gb_est82.*
83: gb_est83.*
84: gb_est84.*
85: gb_est85.*
86: gb_est86.*
87: gb_est87.*
88: gb_est88.*
89: gb_est89.*
90: gb_est90.*
91: gb_est91.*
92: gb_est92.*
93: gb_est93.*
94: gb_est94.*
95: gb_est95.*
96: gb_est96.*
97: gb_est97.*
98: gb_est98.*
99: gb_est99.*
100: gb_est100.*
101: gb_est101.*
102: gb_est102.*
103: gb_est103.*
104: gb_est104.*
105: gb_est105.*
106: gb_est106.*
107: gb_est107.*
108: gb_est108.*
109: gb_est109.*
110: gb_est110.*
111: gb_est111.*
112: gb_est112.*
113: gb_est113.*
114: gb_est114.*
115: gb_est115.*
116: gb_est116.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_esthum29.*
64: em_esthum30.*
65: em_esthum31.*
66: em_esthum32.*
67: em_esthum33.*
68: em_esthum34.*
69: em_esthum35.*
70: em_esthum36.*
71: em_esthum37.*
72: em_esthum38.*
73: em_esthum39.*
74: em_esthum40.*
75: em_esthum41.*
76: em_esthum42.*
77: em_esthum43.*
78: em_esthum44.*
79: em_esthum45.*
80: em_esthum46.*
81: em_esthum47.*
82: em_esthum48.*
83: em_esthum49.*
84: em_esthum50.*
85: em_esthum51.*
86: em_esthum52.*
87: em_esthum53.*
88: em_esthum54.*
89: em_esthum55.*
90: em_esthum56.*
91: em_esthum57.*
92: em_esthum58.*
93: em_esthum59.*
94: em_esthum60.*
95: em_esthum61.*
96: em_esthum62.*
97: em_esthum63.*
98: em_esthum64.*
99: em_esthum65.*
100: em_esthum66.*
101: em_esthum67.*
102: em_esthum68.*
103: em_esthum69.*
104: em_esthum70.*
105: em_esthum71.*
106: em_esthum72.*
107: em_esthum73.*
108: em_esthum74.*
109: em_esthum75.*
110: em_esthum76.*
111: em_esthum77.*
112: em_esthum78.*
113: em_esthum79.*
114: em_esthum80.*
115: em_esthum81.*
116: em_esthum82.*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1333.6	68.5	1354	192	AK005947	Mus musculus AK005947
2	443.2	22.8	871	175	BG291400	BG291400 602387272777
3	418.6	21.3	962	165	BE260911	BE260911 6011538355
4	413.6	21.3	946	139	BE741369	BE741369 601159422277
5	392.4	20.2	544	165	BE253049	BE253049 6011169553
6	336.2	17.3	636	248	AZ720215	AZ720215 RPTC1-24-1.1
7	325.8	16.7	904	141	BE868867	BE868867 6014447277
8	309.6	15.9	799	172	BG026071	BG026071 6022922653
9	297.6	15.3	907	141	BE867810	BE867810 6014433779
10	254.8	13.6	525	31	AV610897	AV610897 AV610897
11	268.8	13.3	669	147	BF110650	BF110650 6018946050
12	249.8	12.8	352	6	AA351060	AA351060 EST58647
13	224.6	11.5	239	123	BB018959	BB018959 BB018959
14	218.2	11.2	616	153	BG392671	BG392671 6024109201
15	206.6	10.6	544	144	BF076210	BF076210 225720 MA
16	193	9.9	245	27	AV270183	AV270183 AV270183
17	189.6	9.7	425	109	AY040695	AY040695 AY040695
18	187	9.6	474	104	A1954626	A1954626 WQ34C08.x
19	183.8	9.4	457	174	BG150572	BG150572 7K02H09.x
20	177.2	9.1	608	175	BG289958	BG289958 6023813795
21	157	8.1	673	154	BG283797	BG283797 6024075121
22	152.6	7.8	679	154	BG483779	BG483779 6025034255
23	151.4	7.8	219	1	AA063239	AA063239 zm02f01.s
24	146.2	7.5	291	6	AA352196	AA352196 EST60152
25	146.2	7.5	531	122	AW666801	AW666801 EST378875
26	122	6.3	607	116	AM649292	AM649292 hc8Og07.x
27	119.6	6.2	368	119	AME75539	AME75539 bd44G06.y
28	117.6	6.0	667	230	AO584348	AO584348 RPTC1-11-4
29	98	5.0	366	5	AA300063	AA300063 EST12698
30	94.8	4.9	257	13	AA919081	AA919081 o178C04.S
31	80	4.1	618	7	AA414516	AA414516 vd09E01.S
32	70.6	3.6	617	138	BE680653	BE680653 dF83C04.y
33	64.4	3.3	597	115	AMA20637	AMA20637 fJ84D09.x
34	62	3.1	928	218	BE678161	BE678161 dF83C04.x
35	59.6	3.2	565	139	CNS0091P	AL053013 Drosophila
36	56.4	2.9	311	5	AA300064	AA300064 EST12699
37	56.4	2.9	716	155	BG362954	BG362954 602381118
38	52.6	2.7	932	219	CNS0072Q	AL056742 Drosophila
39	52.4	2.7	1415	76	AW727479	AW727479 GA-Ea001
40			925	219	CNS0091P	AL053013 Drosophila
41	50.8	2.6	1101	219	CNS010SY	AL108460 Drosophila
42	50.6	2.6	776	219	CNS010RY	AL10993352 Drosophila
43	50.6	2.6	908	219	CNS006B4	AL064031 Drosophila
44	50.4	2.6	935	219	CNS006XK	AL066051 Drosophila
45	49.6	2.5	440	23	A1671575	A1671575 wd33h03.x

ALIGNMENTS

JOURNAL
REFERENCE
AUTHORS

Methods Enzymol. 303, 19-44 (1999)

TITLE

2 (sites)

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Carninci,P.,Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

TITLE

3 (sites)

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Komoto,H., Akiyama,J., Nishi,R., Kitsuami,T., Tashiro,H., Itoh,M., Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishio,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Okazaki,Y., Muramatsu,M., Inoue,T. and Hayashizaki,Y.

Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

TITLE

4 (sites)

JOURNAL
MEDLINE
REFERENCE
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 1354)

JOURNAL
REFERENCE
AUTHORS

Aadachi,U., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Furukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirokawa,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawaji,J., Koimori,Y., Konno,H., Koude,M., Koya,S., Kurihara,C., Matsuura,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and genome science laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to

FEATURES

source

```

/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
<1..1252
/note="putative"
CDS

```


Query Match	Best Local Similarity	Matches	Score	DB	Length
725 GCACGATGATAGGTGAGTGTAGTGAAGAGGCGCTGTCTTGGCACAGACTCTGCAGCTCT 784	79.4%	508; Conservative	413.6; DB 139; Length 946; Pred. No. 1e-102; Mismatches 119; Indels 13; Gaps 1.	13	GCACAAGAGATTAAGTGTGTGTGTGTGAAGAGGCGCCACACCTCCACAGAGAGCCCCAGACTCT 72
785 TTACACTCAAGATCCGGGTCCGGGCTGACCTAGTATGTGAGACTGCCCTGTCAAGATGTGGAGC 844	79.4%	508; Conservative	413.6; DB 139; Length 946; Pred. No. 1e-102; Mismatches 119; Indels 13; Gaps 1.	13	GCACAAGAGATTAAGTGTGTGTGTGTGAAGAGGCGCCACACCTCCACAGAGAGCCCCAGACTCT 72
73 TCCCACTCAAAATCCGTTGTCGGGCTGTGACCTGTGTAGATTTGCCCTCAAGATGTGGAGC 132	79.4%	508; Conservative	413.6; DB 139; Length 946; Pred. No. 1e-102; Mismatches 119; Indels 13; Gaps 1.	13	GCACAAGAGATTAAGTGTGTGTGTGTGAAGAGGCGCCACACCTCCACAGAGAGCCCCAGACTCT 72

OY	845	CCCTTCACAATGTGGTSGAATCAACATGGCCAAATCAACTTTGGGGGTGTCGCCAACAGATTTC	904
Dd	133	CCCTCACAAGTGTGGTGAACACATGGCACACCACCTTGGGGTGTCCCAAGAAGGATTC	192
OY	905	TTTTCTTTTTGGAGAGAGTAACATGTCCTCCACTGCGCACCCCAGTAAACCTTAAGCTTG	964
Dd	193	TTTTGCTTTTTGGAGAGACAGAGCATATCACTACTGCGCACTCCACAGACCCATAAGCTCG	252
OY	965	GAGTGGCTGACATCAATTATTTGTGTGGTGTCTAGACAGACTCTTCAGAGGCCACAGACAT	1024
Dd	253	GAGTGGCTGACATCAATTACTGTGTGGTGTCTAGACAAATTTCTCCAGAGGCCACAGACAGT	312
OY	1025	CCGAGAGACTCCGGGTCCGGGTGCAGAGGGAGAGAAAACCAACATGTTGGAGATCTCAC	1084
Dd	313	CCCAACAGCTCCAGTCCGGGTGCAGAGGAAAAAGAAAACCAACATGGAATCTTCAC	372
OY	1085	TGTCCTCGAATTCCTCTTAAGGTTCTCATGTCACTATGAGAGAACCATGGACTCT	1144
Dd	373	TGTCCTCGAATTCCTCTTAAGGTTCTCATGTCACTATGAGAGAACCATGGACTGT	432
OY	1145	CMGACACAAAGCTCCCTCTCTTTGATGGAGCAAAGCTTTCAGGAAGAGCTGCAG	1204
Dd	433	CGGAGCGAAGCTCTCTCTCTTTGATGGAGCAAAGCTTTCAGGAAGAGAGCTGCAG	492
OY	1205	CTGATCTGGGGCTGGAAATCCGAGATCTCAGTCAGAAAGTGTGGGGCTGAAGCTTCACCTG	1264
Dd	493	CTGATCTGGGGCTGAATCTGGGGACCTCATTTGAGGTGTGGGGCTGACACCCACCTCCT	552
OY	1265	TTTGGACCCAAGCCAAACATGGAGACATTAAGCTCCCAATTTATTTATTTGTGATTTTC	1324
Dd	553	GTTTTACGCGCCCGAGCTCACTTGGGGAGAAATG-----ACTTTCCTCTTTT	599
OY	1325	GCCCATTAAGGCTTACAGAAATCAATTAAGACTTGT	1364
Dd	600	GCCCCATAAGGCTTAGCATTAAGCTGAGAGTAACTTATCT	639

RESULT	5		
LOCUS	BE253049	544 bp	mRNA
DEFINITION	601116953r1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3357421 5',	EST	13-JUL-2000
ACCESSION	BE253049		
VERSION	BE253049.1	GI:9123199	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
TITLE	1 (bases 1 to 544)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs@email.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	GenBank Entry: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: Image.lnl.gov plate: LNCMI64 row: j column: 14		
	High quality sequence stop: 542.		
FEATURES	Location/Qualifiers		
source	1..544		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3357421"		
	/clone_id="NIH_MGC_16"		
	/tissue_type="retinoblastoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:		

Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Comment
BE868967	1	GI:10317730		Homo sapiens human.					
BE868967.1	EST.			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
1 (bases 1 to 904)									
NIH-MGC	http://mgc.nci.nih.gov/								
Unpublished (1999)									
Contact: Robert Strausberg, Ph.D.									
Email: cgapbs-remail.nih.gov									
Tissue Procurement: ATCC									
cDNA Library Preparation: Life Technologies, Inc.									
DNA Sequencing by: Incyte Genomics, Inc.									
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:									
http://limage.llnl.gov									
plate: lImage9565	row: e	column: 08							
High quality sequence stop: 589.									
Location/Qualifiers									
1..904									
/organism="Homo sapiens"									
/db_xref="taxon:9606"									
/clone="IMAGE:3848815"									
/clone_lib="NIH-MGC-65"									
/tissue_type="adenocarcinoma"									
/lab_host="DH10B (phage-resistant)"									
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.8 kb. Library constructed by Life Technologies."									
BASE COUNT	217 a	235 c	251 g	201 t					
ORIGIN									
Query Match	16.7%	Score 325.8;	DB 141;	Length 904;					
Best Local Similarity	82.7%;	Pred. No. 1.6e-78;							
Matches 444;	Conservative	0;	Mismatches 87;	Indels 6;	Gaps 6				
774	CTCTGCACCTTTTACACTCAAGATCCGGGTCCGGGCTGACCTTAGTACAGACTGCTGTGAC	833							
31	CCCCGACCTCTTCCCACTTCAAGATCGTTTCCGGGCTGACCTGATGATCA-TCGCCCCTCAG	89							
834	GATGTGCGAGCCCTTCAGATGTGTGTGATGCACATGAGGCAATCATTTGGGGTGTCTCC	893							
90	GATGTGCGAGCCCTTCAGATGTGTGTGATGCACATGAGGCAATCATTTGGGGTGTCTCC	149							
894	AAACAGGATTTCTTTGCTTTTGGAGAGAGTGAACGTCTCTACTGCGCACCCCTAGTAC	953							
150	AAGCAGGATTCCTTTTTC-TTTTGGAAGAGACAGAGCATATCACTATGCGACCTCCAGAGC	208							
954	CCATAAGCTTGGAGTGGGTGACATCATTTGTTGTGTTGGGTCTGTAGCAAGCTCTTCAGAGGC	1013							
209	CCATAAGCTTGGAGTGGGTGACATCATTTGTTGTGTTGGGTCTGTAGCAAGCTCTTCAGAGGC	268							
1014	CACAGAGACATCCACGAGAGCTCCGGCTCCGGGTGCAAGGGAGAGAGAAACACAGATGTT	1073							
269	CACAGAGAGCTCCCAACAGCTCCAGCTCCGGGTGCAAGGGAGAGAGAAACAGACACT	327							
1074	GGAGATCTACAGTCTTCGATTTCTCTCTTAAGGTTCTCATGTCAACATATAGAGAAC	1133							
328	GGAGATCTACAGTCTTCGATTTCTCTCTTAAGGTTCTCATGTCAACATATAGAGAAC	386							
1134	CATGGAGCTCTTGAGACACAAAGCTCTCTCTTTTGATGGAGACAAGCTTTTCAGACAA	1193							
387	CATGGAGCTCTTGAGACACAAAGCTCTCTCTTTTGATGGAGACAAGCTTTTCAGACAG	446							
1194	GGAGCTCCACAGTACTGTGGCCCTGGAAATCCGAGATCTCATGAAAGTCTGGGGCTGAAG	1253							
447	GGAGCTCCACAGTACTGTGGCCCTGGAAATCCGAGATCTCATGAAAGTCTGGGGCTGAAG	505							

QY	1254	CHCTCAACCCGTTGGAGGCAAGCAACACATGAGACATAGAGCCCAATTTAT	1310
Db	506	ACCCATCACTGTT-TCAGCGCCAGCCTGAGCTTGGAGAAATGACTTTCCTTTT	561
RESULT	8		
LOCUS	BC026071	799 bp	EST 24-JAN-2001
DEFINITION	602292263p1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4387022 5', mRNA sequence.		
ACCESSION	BC026071		
VERSION	BC026071.1	GI:12413308	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 799)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://limage.llnl.gov Plate: LLM10070 row: n column: 15 High quality sequence stop: 529. Location/Qualifiers 1. 799 /organism="Homo sapiens" /db_xref="taxon:9606" /IMAGE:4387022 /clone_1db="NIH_MGC_86" /clone_1db="NIH_MGC_86" /tissue_type="osteosarcoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: bone; Vector: PCMV-SF0RN6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.53 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."		
BASE COUNT	210 a 239 c 210 g 140 t		
ORIGIN			
Query Match	15.9%	Score 309.6; DB 172; Length 799;	
Best Local Similarity	74.3%	Pred. No. 4,4e-74;	
Matches	547; Conservative	0; Mismatches 154; Indels 35; Gaps 11.	
QY	641	AGAACTAAGGAGAGTGAACAAGCGCTTCTCAAGATCTCGCTCGCCAGAGCCCAAGC	700
Db	2	AGAAAGTTAAGTAGAGTGAGAACAAAGCGCTCAGAGATCTCGTCTGCTGAGAGCCCAAGC	61
QY	701	AGCACCAGAGTCCAGCCCTTCAGAGACACAGATGATGAGTGGTCTAGTGGAGAGGCGCTG	760
Db	62	CACCTTCAGAGGTGAAGAGCAACAGGGCCAAAGAGATGAACTGGCTTGGTGGAAAGGCCCA	121
QY	761	TCCTTGCCACAGAGCTCTGACACTCTTACACTCAAGATCGGTCGGGCGGAGCTGACCTAGTGA	820
Db	182	GATTTGCCCTCAGAGATGTGGAGAGCCCAATGACAGAGTGTGGACACATGGCCACCCAC	241
QY	821	GACTGCTGTGTCAGAGTGTGAGAGCCGCTTCAGAAATGTGTGGATTCACATGGCCAAATCAT	879
Db	880	CTTTGGGGTCTCCCAACAGATTTCTTTTGGCTTTTGGAGAGATGATGCTCTCTACT	939
QY	242	CTTTGGGGTGTCCCAAGAGAGATCTTTTTCGTTTGGAGACAGAGCTATCACCCTACT	299

Query Match	15.3%;	Score 297.6;	DB 141;	Length 907;
Best Local Similarity	72.1%;	Pred. No. 9.2e-71;		
Matches 528;	Conservative 0;	Mismatches 184;	Indels 20;	Gaps 10;
QY 324 GCTGCTGATCCCGAGAGGGCGCGGTGTGCCAG-TGTACTCCGGGAAGGTACGAGACA 382				
Db 12 GGTGCTGATCCGGGGGAGGGCGCGCTGGTTCGGGATCTACTCGGGGAAGGTAAACACA 71				
QY 383 GCGTCACCCCTCATTTCCACATATTAATTCATCCCTCTTGAACCTGCGCCCTTCAGAGCCCTGAAG 442				
Db 72 GCGCTTCGCGCTTATCCACATGATCTATCTCCCTCTGTAACCTTACCCTCCAGGGGATGAGG 131				
QY 443 ATGAGCGAGATCTGCACAATTCCTGGCAGTCTCCCTCTGAGAGATGATCCCTGCGCTTACG 502				
Db 132 AAGAGCGAGAGCTGGCAGATTC--GATGGCTCTACCATGAGAGGCTGCCCATCAGAG 188				
QY 503 GTTCTCCCTGGAGAAAGAGCTCAGAAAGAAAGTGTGAGAAAGAGAAAGAAATGTGAG 562				
Db 189 GCTCTCCCTGGAGCAAGAGCTGAGAGACTAAG--GATTAAGAGAGAGAGAGAGAGAGAG 245				
QY 563 AGTTTCGGAGACAGACATCTCTCCCTTGGCCCAACCTTGCAGAGAGAGAGAGAGAGAGAG 622				
Db 246 AGTTCTCGATCTGGACAACTCTCTCTGTCCCACTTCAACAGAGAGAGAGAGAGAGAGAG 305				
QY 623 AGCATACGGAGGCGCTCCAGAGCTTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 682				
Db 306 CGCATACCTGGGCACTCAAGAAAGTAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365				
QY 683 CCTGCTGAGGCCCCAGAGCAGCAGAGTCCAGCCCTTTCAGAGCAGAGAGAGAGAGAGAG 742				
Db 366 CCTGTCTAGGCCCCAGAGCAGCAGAGTCCAGAGGCTCAAGAGCAAGAGAGAGAGAGAGAG 425				
QY 743 TCCGTGTGAAGGGGCTCTTGGCCACAGAGGTCTCGACTCTTTCACATCAAGATCCGGT 802				
Db 426 TCTTGTGTGGCAAGGGCCACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 483				
QY 803 GCGGCGGCTGACCTTACTGAGAGCTGCTGTCCAGAGATGTGAGAGGCCCCCTTCAGAGATGTGTGG 862				
Db 484 GCGGCGGCTGAG--TGGTCAGATGCCCCCTCCAGAGATGTGAGAGGCCCCCTTCAGAGATGTGTGG 541				
QY 863 ATCAGATGCGCCATCATCTTTGGGCTGTCTCCAAACAGAGATTTCTTTTGTGTGTGAGAGA 922				
Db 542 AACCAATATGGC--ACCACTTGTGGGTGTCCCAAGCAGAGATCTTTT--GCTTACGAGAGAGA 596				
QY 923 GTGACATCTCTCTCTCTCTGCGCAC-CCCTAGTACCTTAAGAGCTTGGAGTGGGTGACATCTTT 981				
Db 597 CAGACATCTCTCTCTCTCTGCGCACCAAGAGCGGCTTAAGAGCGGGGAGGTGGTACATCTAT 656				
QY 982 GA-TTGT 1039				
Db 657 GACGGGGGGGTGATTAACCAAGATCTCTCAGAGAGGAGCA-CGAGAGCTCCCAAGGTTCAGC 715				
QY 1040 TCCGGGTGAGG 1051				
Db 716 TCCGGCTGACAG 727				

REFERENCE 1 (bases 1 to 525)
 AUTHORS Sugimoto,Y., Hirotsune,S., Takasuga,A., Itoh,R., Ujihazono,A. and Suzuki,H.
 TITLE Bovine CDNA sequencing
 JOURNAL Unpublished (2000)
 COMMENT Contact: Yoshikazu Sugimoto
 Animal Genetics Division
 Shikawa Institute of Animal Genetics
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
 Tel: 81-248-25-5641
 Fax: 81-248-25-5725
 Email: kazusugi@cocoa.ocn.ne.jp
 Single pass sequencing.
 This clone was obtained from a polyA-deleted cDNA library.

FEATURES
 Location/Qualifiers
 1..525
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone="B1LP035D02"
 /clone_lib="Bos taurus lung fetus"
 /tissue_type="lung"
 /dev_stage="fetus"
 /lab_host="DH10B"
 /note="Vector: pZ11; Site_1: SalI; Site_2: NotI; Poly A was deleted from a NotI site"

BASE COUNT 138 a 144 c 147 g 94 t 2 others
 ORIGIN

Query Match 13.6%; Score 264.2; DB 31; Length 525;
 Best Local Similarity 74.0%; Pred. No. 1.2e-61;
 Matches 364; Conservative 0; Mismatches 119; Indels 9; Gaps 2;

346 CCGGTGTCCTCCATGTCCTCCGGAGGCTACAGACGACCTCAACCTCAGATCAAT 405
 1 CCGGGCTTCCTGATCTGCTGAGAGGAGTAAAGCCTCACTTCCAGATAAC 60
 406 TCATCCCTCTTGAACAGTGTCCCTTCAGAGCTGAAGATGAGCAGATCTGACAAATTC 465
 61 GTTCCTCTGAAATATATGTCCCCAGAGCTGAAGAGGCAAGATGTGAGAGATGCC 120
 466 GGCAGTTCTCCCTGAGATGATGCCCTTCAGGTTCTCCAGAGAAAGAAAGCTC 525
 121 AGCAG--TCCCATGCAAGTGTTCCTCATGTCCAGTTCCCTCGAAGAAAGCT- 176
 526 AGAAGAGTGTAGAAAGAAAGAAATGGAAGATTCCGGACCCAGACATCTCT 585
 177 -----GAGGAGTAAGATGAGAGAGAAAGAAAGATGTACTGATCAGACACTCG 231
 586 CTTTGGCCCAACCTTCGTAAGAACAAAGACATAAGGAGGCTCCAGAAAG 645
 232 CTTTGGCCCTCAGCTCGCAGCAGTACCAAGAGAGAGATACAGGATCCGGAAG 291
 646 CTAAAGGAGTGAACAAGCTCTCCAGATCTCCGCTCGCTGAGCCCAAGCAGAC 705
 292 TTAAAGGAGGTGAACAAGCTCTCCAGATCTCCGCTCGCTGATCCCAAGCAATCC 351
 706 CAGAGTCCACCTCTCAGACAGAGATGATGAGTGTCTTACTAGTGAAGGCTCTCTTG 765
 352 CAAGGCCCAACCACTGAGCCCAAGAGATGATGATGCTTACTAGTGGGGCCCAAGNCTC 411
 766 CCACAGAGCTCTGACTCTTACACTCAAGATCCGATGCCGCTGACCTAGTGAAGACTG 825
 412 CCAGAGAGCTCTGACTCTTCCGCACTCAAAATCCGGTCCGAGACTGACTGGTCAAGTTG 471
 826 CCTGTGAGATG 837
 472 CCCATCAGATG 483

RESULT 11
 BF310650

LOCUS BF310650 669 bp mRNA EST 21-NOV-2000
 DEFINITION 601894605F2 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4124232 5',
 mRNA sequence.
 ACCESSION BF310650
 VERSION BF310650.1 GI:11258236
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 669)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LHCMI010 row: 1 column: 01
 High quality sequence stop: 600.
 Location/Qualifiers
 1..669
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4124232"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 174 a 191 c 184 g 120 t
 ORIGIN

Query Match 13.3%; Score 258.8; DB 147; Length 669;
 Best Local Similarity 70.4%; Pred. No. 4.1e-60;
 Matches 432; Conservative 0; Mismatches 172; Indels 10; Gaps 6;

416 TGAACCTGTGCTTCAGAGCTGAGAGTGAAGCAGATCTGACAAATTCTGCAGTTCTC 475
 2 TGAACCTGTACCTCCAG--GGATGAGAGAGGACAGCTGCAGATTC---GAGTGGTC 57
 476 CCTCTGAGATGATGCCCTTCCTCAGGTTCTCCCGAGAAAGAGCTCAGAAAGACT 535
 58 TCTACCATGAGGGCTCCCATCACAGGCTCTCCGGAAGACAAAGCTGAGACTAAG- 116
 536 GTGAGAAGAGAAAGAAATGAGAGTTCGAGACAGACATCTGCTTGGCCCC 595
 117 --GATAAAGAGAGAGAAAGAAACAGAGTTTGTGATCTGACAACTCTCCTCTGCC 174
 596 AACCTTGTCAAGGAACAAAGCAGAAAGCATAGGAGGCTCCAGAAAGCTAAGGAAG 655
 175 CACCTTACCAACAGAACCAAGAGCAGAACGATCTGCGGCACTCAGACAGATTAGTGAAG 234
 656 TGAACAAGCTCTCCAAAGATCTCCCTGCTGAGGCCCAAGCAGACAGACAGCTCAG 715
 235 TGAACAAGCGCTTCAGAGATCTCGTTCTGTGAGCCCCAAGCCACCTCAGAGGCTAAG 294
 716 CCTTCAGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 774
 295 AGCAACAGGGCCCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 354
 775 TCTGACTCTTTACACTCAAGATCCGGTCCGGCTGACCTAGTGAAGACTGCTGTCAAG 834

FEATURES

VERSION BF076210.1 GI:10869849
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 544)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
W.W. and Keele,J.W.

TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCA
BACKWARD: GTTTCAGTCAGCAGC
Plate: 86 row: G column: 5
Seq primer: ATTTCAGTGCACATATAG.

FEATURES
Source Location/Qualifiers
1..544

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_id="MARC 280Y"
/tissue_type="pooled"
/lab_host="DH10B"

/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

BASE COUNT 106 a 156 c 147 g 135 t
ORIGIN

Query Match 10.6%; Score 206.6; DB 144; Length 544;
Best Local Similarity 79.8%; Pred. No. 9.1e-46;
Matches 256; Conservative 0; Mismatches 64; Indels 1; Gaps 1;

QY 978 CATTGATTGTGTGCTAGCAGCTCTTCAGAGGCCACAGACATCCAGAGCTCCG 1037
DB 35 CACGAGCTGTGTGTAGCAGAGTCTCCGAGGCTGCAGAGAGCTGCCACTGCTCCA 94

QY 1038 GCTCGGGGTGAGGGGAAGGAACAACAGATGTGAGATCTCACTGTCTCCTGATTC 1097
DB 95 ACTTCGGGTCCAGGGAAGGAAGCAACAGAGCTGGAAGTCTCGCTCCCTCCGATTC 154

QY 1098 TCCTTAAGTGTCTCATGTACACTATGAGAGCCATGGAGCTCTGTGACACAAGCT 1157
DB 155 TCCTCTCAAGACCTCATGTCCGCTAGAGAGGAGCTATGAGACTCTCCGCCACAAGCT 214

QY 1158 CTCCTTTCTTTGATGAGCAAAAGCTTTTCAGGCAAGAGCTGCCAGCTGATCTGGGCT 1217
DB 215 CTCCTTCTTTTGTGATGAGCAAGCTTTCAGGCAAGAGCTGCCAGCTGACCTGGGAAT 274

QY 1218 GGAATCCGAGATCTCATTCAGAGCTGGGGCTG-AAAGCTCAACCTGTTCGAGCAAA 1276
DB 275 GGAATCTGGGAGCTCATTCGAGGCTGAGGCTGAACCCCACTCCTGTTGGAGGCCA 334

QY 1277 GCCAAGACATGAGACAATAG 1297
DB 335 GCCTGAGCTTGGGGAATAG 355

Search completed: September 11, 2001, 12:52:44
Job time: 4256 sec

KM	hyperplasia; ds.
OS	Homo sapiens.
PN	W09964576-A2.
PD	16-DEC-1999.
XX	
PF	09-JUN-1999; 99WO-IB01062.
PR	10-JUN-1998; 98US-0088801.
PA	(FARB) BAYER CORP.
PI	Endege WO, Steimann KE, Astle JH, Burgess CC, Bushnell SE;
PI	Carroll E, Calino TU, Dertl A, Ford DM, Lewis ME, Monahan JE;
DR	Schlegel R;
XX	
XX	WPI; 2000-087220/07.
PT	
PT	Novel nucleic acids, used to develop products for the diagnosis and
PT	treatment of disorders involving unwanted cell proliferation,
PT	particularly cancers, especially colon cancer .
XX	
PS	Claim 1; Page 160; 469pp: English.
CC	
CC	AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC	the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC	cDNA clones can be used to generate antisense oligonucleotides which
CC	can be used for antisense therapy. Methods and products from the present
CC	invention can be used for identifying and/or classifying cancerous cells
CC	present in a human tumour, particularly in solid tumours, e.g.
CC	carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC	can be used for developing agents for the diagnosis and treatment of
CC	disorders involving unwanted cell proliferation, such as neoplasia,
CC	dysplasia or hyperplasia.
XX	
XX	
SQ	Sequence 705 BP; 190 A; 134 C; 194 G; 157 T; 30 other;
<hr/>	
Query Match	8.6%; Score 167.6; DB 21; Length 705;
Best Local Similarity	86.4%; Pred. No. 4.7e-37;
Matches 185; Conservative	0; Mismatches 29; Indels 0; Gaps 0
OY	779 GACTCTTACACTCAAGATCGGTCGGGGGTGCAGCTAGTAGACTCCTGTAGAGATGT 838
Db	214 GCCTCTTCCCACTCAAAATCCCTGCCGGGCTGACACTGTGAGATTGGCCCCCTCAGCATGT 155
OY	839 CGGAGCCCCCTAGAGATGGTGAGTACATGGCCAATATCTTTGGGGGTGTCTCCAAACA 898
Db	154 CGGAGCCCCCTGAGAAGTGTGTGAGACACATGGCCACCACCTTGGGGTGTCCCACAGA 95
OY	899 GGATTCCTTTTGGTTTTTGGAGAGAGTAACACTGTCTCTACTGCCAACCCCTAGTACCCTAA 958
Db	94 GGATTCCTTTTGGTTTTTGGAGAGACAGACACTTTCACCTACTGCCAATCCAGAGACCTAA 35
OY	959 AGCTTGSAGTGGCTGACATCATATTGAATTTGTGTGT 992
Db	34 AGCTGGAGTGGCTGACATCATATTGACTGTGTGT 1
<hr/>	
RESULT	7
AAQ73500	
ID	AAQ73500 standard; DNA; 8438 BP.
XX	
AC	AAQ73500;
XX	
DT	15-MAY-1995 (first entry)
XX	
XX	DNA encoding Pseudorabies virus large latency transcript.
XX	
Pseudorabies virus; PRV; ILV; large latency transcript;	
attenuated virus; vaccine; early protein 0; Ep0; HSV-1 ICP0;	

[illegible]

XX	Sequence	114955 BP;	6071 A;	29417 C;	36712 G;	21328 T;	21427 other;
SO	Query Match	2.2%;	Score 42.8;	DB 20;	Length 114955;		
	Best Local Similarity	32.5%;	Pred. No. 0.75;				
	Matches 111;	Conservative 38;	Mismatches 193;	Indels 0;	Gaps		
OY	8 GGGAGATGGCGGAACCACTAGAGGGAGACTGGTCCGAGGTCCCGGGTGGCCGAGGGCTTC	67					
DB	104714 GNNHHNNNGCGCGCCVGGCCVGGCGNNHHNNNSGGCGCGCCVGGCCVGGCGGNNHHNNNSC	104655					
OY	68 GGAGAGCCCGAGGGCGCCGTCGGCGGTGCTCTCGCGCCGCGGACGTCTCCGGCTAGGCTCA	127					
DB	104654 GGCGCGCCCGVGGCCVGGCGNNHHNNNSCGCGCGCCCGCCCGCCVGGCCVGGCGNNHHNNNSGGCGGC	104595					
OY	128 TTCCAGACACCGCTCTTGAGCTTGTGACTTGTGACTAGTGACAGCAGCAGAGAGTCTTGGAAGTGC	187					
DB	104594 GCGCCCGVGGCCVGGCGNNHHNNNSCGCGCGCGCCCGCCVGGCCVGGCGNNHHNNNSCCCGCGG	104535					
OY	188 CAGACCCGAGTGAAGTGTGCGGCTGCGCGCCGCTCCCGCGCGCGCGGTGAACCTGAGCAGACACA	247					
DB	104534 GCGCCCGVGGCCVGGCGNNHHNNNSCGCGCGCGCGCCCGCCVGGCCVGGCGNNHHNNNSGGCC	104475					
OY	248 GCGACAGTGAAGTGAAGAGGGCGCGGAGGGCGCTGCGGAGAGCCCGCGCTACATTGGTGC	307					
DB	104474 GCGCGCGCGCCVGGCCVGGCGNNHHNNNSCGCGCGCGCGCGCCCGCCVGGCCVGGCGNNHN	104415					
OY	308 GACGCGCGCGCGCGCGCTGCTGATGCCGAGAGGCGCGCGG	349					
DB	104414 NNSCGCGCGCGCGCGCGCCVGGCCVGGCGNNHHNNNSCCG	104373					
RESULT 10							
XX	AAK53491						
AC	AAK53491 standard; DNA: 114955 BP.						
XX	AAK53491:						
DT	05-JUL-1999 (first entry)						
XX							
DE	Human adenosine A1 receptor antisense oligonucleotide fragment.						
XX							
KM	Antisense oligonucleotide; multiple target; antisense treatment;						
KM	Impaired respiration; inflammation; lung disease;						
KM	pulmonary vasoconstriction; inflammation; allergic rhinitis;						
KM	acute asthma; allergy; asthma; impaired respiration;						
KM	respiratory distress syndrome; pain; cystic fibrosis;						
KM	pulmonary hypertension; pulmonary vasoconstriction; emphysema;						
KM	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;						
KM	colon cancer; breast cancer; lung cancer; pancreatic cancer;						
KM	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;						
KX	prostate cancer; ss.						
OS	Synthetic.						
XX							
PM	WO913886-A1.						
XX							
PD	25-MAR-1999.						
XX							
PF	17-SEP-1998; 98WO-US19419.						
XX							
PR	09-JUN-1998; 98US-0093972.						
PR	17-SEP-1997; 97US-0059160.						
XX							
PA	(UYEC-) UNIV EAST CAROLINA.						
XX							
PI	Nyce JW;						
XX							
DR	WPI: 1999-229400/19.						
XX							
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary						
PT	vasoconstriction						


```

XX Key Location/Qualifiers
FH misc_feature 1..1078
FT misc_feature /tag= a
FT misc_feature /label= St B12.3
FT misc_feature /label= St B12.5
FT misc_feature /tag= b
FT misc_feature /label= St B12.5
XX
XX MO9214840-A.
XX
XX 03-SEP-1992.
XX
XX 13-FEB-1992; 92MO-FR00145.
XX
XX 13-FEB-1991; 91FR-0001684.
XX
XX (HEID/) HEIDET S.
XX (INRM ) INSERM INST NAT SANTE & RECH MED.
XX
XX Heliz D, Mandel J, Oberle I, Rousseau F, Vincent A;
XX Mandel JL;
XX
XX WPI; 1992-316192/38.
XX
XX Nucleic acid fragment of X chromosome region - for diagnosis of
XX fragile X syndrome and related mental retardation
XX
XX Claim 8; Fig 8; 53pp; French.
XX
XX Probe 9B12.3 (or "St B12.3") is used to detect anomalies at the
XX fragile X site after digestion of genomic DNA by EcoRI or HindIII.
XX This and other probes, including St B12.5, which can be used in the
XX diagnosis of Fragile X syndrome, were obtained by subcloning DNA
XX fragments from the 225kb region between probes St677 and Do33. This
XX region is localised to the Xq27-q28 region of the human X chromosome
XX cloned as an artificial yeast chromosome. Specifically, the probes
XX are adjacent to or comprise CpG sites.
XX See also AAQ28291-3 and AAQ28295.
XX
XX Sequence 2183 BP; 520 A; 655 C; 533 G; 460 T; 15 other:

Query Match 2.1%; Score 41.6; DB 13; Length 2183;
Best local Similarity 67.0%; Pred. No. 0.19; Mismatches 29; Indels 0; Gaps 0;
Matches 59; Conservative 0;

QY 243 GGCAGCGAGAGTGCAGTGTGAAGGGGCGCCGAGGGGCTGCGGAGCCCGCGTACATT 302
DB 1750 GGGGGGGGGGGTGCAGGAGGGCGCGCTGCGCAGGGGGCTGCGGAGCGGGCGCGG 1691
QY 303 GGTGCGAGGGGGGGGGGGGGGGGGTCTG 330
DB 1690 GCGGGGGGGGGGGGGGGGGGGGGGG 1663

RESULT 13
AAAX23299/C
ID AAAX23299 standard; DNA; 3656 BP.
XX
XX AAAX23299;
XX
XX 11-JUN-1999 (first entry)
XX
XX Human SPA-1 cDNA.
XX
XX SPA-1; tumour-suppressing; low molecular weight G protein activation;
XX treatment; gene therapy; cancer; human; ds.
XX
XX hmo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 297..3425
FT

```

```

FT /*tag= a
FT /product= "SPA-1"
XX
XX MO910380-A1.
XX
XX 04-MAR-1999.
XX
XX 21-AUG-1998; 98MO-JP03715.
XX
XX 22-AUG-1997; 97JP-0240218.
XX
XX (MINA/) MINATO N.
XX
XX Minato N;
XX
XX WPI; 1999-190588/16.
XX
XX P-PSDB; AAM93489.
XX
XX New human SPA-1 protein - has a strong tumour suppressing capacity
XX Claim 1; Page 29-37; 78pp; Japanese.
XX
XX This invention describes a novel human SPA-1 protein with a strong
XX tumour-suppressing activity and which is capable of activating low
XX molecular weight G proteins. Human SPA-1 is useful for the treatment
XX (including gene therapy) of cancers in humans.
XX
XX Sequence 3656 BP; 667 A; 1257 C; 1161 G; 571 T; 0 other:

Query Match 2.1%; Score 40.2; DB 20; Length 3656;
Best local Similarity 49.4%; Pred. No. 0.63; Mismatches 134; Conservative 0; Mismatches 133; Indels 4; Gaps 1;
Matches 134;

QY 31 GGCAGTGTCCGAGTGTCCCGCGGTGGCCGAGGCGCTCGAGACCCGAGCGCCCGTGGC 90
DB 2538 GGAAGTGGGAGGCGCTCGGGGAGGCTGGGACCGCTGCGCGACACGCGAGGA 2479
QY 91 CGAGTGTCTCGCGCGCGCGAGTCTCGGCTAGGCTCATTCAGACAC---CGTGTGT 146
DB 2478 GCGGCGCGCGCGCGCGCGAGCGCCCGCGCTCGCGGAGTGTGAAGCGCTCACGTCGA 2419
QY 147 GACCTGTGTCAGTGCAGAGCGAGAGAGTCTTGAAGTGCAGACCCAGTAGAGTGC 206
DB 2418 CGAATCCCTGCGGTCCACCTCGAAGGCCAGCGGCGCTTGACCGTGGCGGAGCGCCA 2359
QY 207 GGTGCGCGCGCTCCCGCGCGCGCTAACCCTGAGCAGAGACGAGACAGTGAAG 266
DB 2358 GCTGCGGGTCTGCGACACCGGCTCACAGCTGCAAGCGCGCACCACTGCGCCACGG 2299
QY 267 GCGGCGCGAGGGGCTGCGGAGGCGCGGT 297
DB 2298 CTTGCGCGGGAGACCGTCTGAGAGCGAGCGT 2268

RESULT 14
AAT72781/C
ID AAT72781 standard; cDNA; 4425 BP.
XX
XX AAT72781;
XX
XX 23-SEP-1997 (first entry)
XX
XX Human Interleukin-1 receptor interacting protein cDNA.
XX
XX Interleukin-1 receptor interacting protein; IL1; IL-1; inhibitor;
XX tumour necrosis factor; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 1108..4236
FT /*tag= a

```


Wed Sep 12 06:07:09 2001

us-09-192-611-1.rng

Page 12

[illegible]

Search completed: September 11, 2001, 12:59:14
Job time: 4501 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2001, 11:42:58 ; Search time 72.63 Seconds
(without alignments)
5072.277 Million cell updates/sec

Title: US-09-192-611-1

Perfect score: 1946
Sequence: 1 ACAGTGTGGAGATGGCGGA.....CGCGATTCGGCGACTCGAG 1946

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1946	100.0	1946	2	US-08-755-584-1
2	1946	100.0	1946	3	US-09-192-611-1
3	46.4	2.4	8438	1	US-07-945-283-1
4	44	2.3	1026	4	US-07-751-891B-24
5	44	2.3	4188	4	US-07-751-891B-2
6	44	2.3	4242	4	US-07-705-490-2
7	44	2.3	5222	4	US-07-751-891B-23
8	41.6	2.1	4362	2	US-08-455-073A-1
9	38.6	2.0	1002	2	US-08-890-980-5
10	38.6	2.0	1002	3	US-08-890-979-5
11	38.6	2.0	1002	3	US-09-032-894-5
12	38.6	2.0	1002	3	US-09-032-894-95
13	38.6	2.0	1002	4	US-09-031-626-5
14	38.6	2.0	1002	4	US-09-031-626-95
15	38.6	2.0	246240	2	US-08-724-394A-20
16	38.6	2.0	246240	2	US-08-724-394A-21
17	38.6	2.0	246240	2	US-08-724-394A-22
18	38.4	2.0	7218	1	US-08-232-463-14
19	38.2	2.0	4897	6	5196516-7
20	37.4	1.9	43280	2	US-08-804-227C-1
21	37	1.9	11219	1	US-07-642-734C-1
22	37	1.9	11219	3	US-08-439-009A-1
23	36.2	1.9	2965	3	US-08-906-360-2
24	36	1.8	2219	3	US-08-510-646B-17
25	36	1.8	38506	3	US-09-320-878-19
26	35.8	1.8	44377	2	US-08-804-227C-7
27	35.8	1.8	44377	2	US-08-804-198-1

C	28	35.6	1.8	1931	2	US-09-130-114-2	Sequence 2, Appl
C	29	35.4	1.8	13987	2	US-08-804-227C-13	Sequence 13, Appl
C	30	35.2	1.8	774	3	US-08-956-307B-12	Sequence 12, Appl
C	31	35.2	1.8	778	3	US-08-956-307B-11	Sequence 11, Appl
C	32	35.2	1.8	1788	3	US-08-303-861-1	Sequence 1, Appl
C	33	34.8	1.8	423	1	US-08-470-179-113	Sequence 113, App
C	34	34.8	1.8	1020	2	US-08-475-844-10	Sequence 10, Appl
C	35	34.8	1.8	1020	5	PCT-US95-08429-10	Sequence 10, Appl
C	36	34.8	1.8	8056	3	US-09-136-605-14	Sequence 14, Appl
C	37	34.8	1.8	8082	1	US-08-306-691B-41	Sequence 41, Appl
C	38	34.8	1.8	8082	1	US-08-187-785-1	Sequence 1, Appl
C	39	34.8	1.8	8082	5	PCT-US93-06251-28	Sequence 28, Appl
C	40	34.6	1.8	2091	3	US-08-899-437-22	Sequence 22, Appl
C	41	34.6	1.8	2091	4	US-09-126-121-22	Sequence 22, Appl
C	42	34.6	1.8	2502	3	US-08-899-437-5	Sequence 5, Appl
C	43	34.6	1.8	2502	4	US-09-126-121-5	Sequence 5, Appl
C	44	34.2	1.8	2562	1	US-08-146-421-4	Sequence 4, Appl
C	45	34.2	1.8	12001	1	US-08-458-568A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-755-584-1
; Sequence 1, Application US/08755584
; Patent No. 5858711
; GENERAL INFORMATION:
; APPLICANT: Glimcher, Laurie H.
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. Decont, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUI-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..1248
; US-08-755-584-1

Query Match 100.0%; Score 1946; DB 2; Length 1946;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1946; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ACAGTGTGGAGATGCGGGAACCACTGAGGGACGTGTCCGAGTCCCGGTGCGCA 60
    |||||||
Db 1 ACAGTGTGGAGATGCGGGAACCACTGAGGGACGTGTCCGAGTCCCGGTGCGCA 60
QY 61 GGGGCTCGAGAGACCCGAGGCGCCGCGGTGCGGTCTCGCGCCGCGAGTCCGAGCT 120
    |||||||
Db 61 GGGGCTCGAGAGACCCGAGGCGCCGCGGTGCGGTCTCGCGCCGCGAGTCCGAGCT 120
QY 121 AGGCTCATTTCCAGACACCGCTGTGTGTGAGCTTGTGACAGACGACGAAGAGTCTTG 180
    |||||||
Db 121 AGGCTCATTTCCAGACACCGCTGTGTGTGAGCTTGTGACAGACGACGAAGAGTCTTG 180
QY 181 GAAGTGGACAAACCACTGATGAGGTGCGGTGCGCGCTCCCGCGCGCTAAACCTGAG 240
    |||||||
Db 181 GAAGTGGACAAACCACTGATGAGGTGCGGTGCGCGCTCCCGCGCGCTAAACCTGAG 240
QY 241 CAGGACAGCAGACAGTGCAGTGAAGGGGCGCGAGGGGCGCTCGGGAGGCGCGCGTACA 300
    |||||||
Db 241 CAGGACAGCAGACAGTGCAGTGAAGGGGCGCGAGGGGCGCTCGGGAGGCGCGCGTACA 300
QY 301 TTGGTGGACGCGCGCGCGCGCGCTGTGTGATCCGGAGAGGCGCGGTGTCCAGTG 360
    |||||||
Db 301 TTGGTGGACGCGCGCGCGCGCGCTGTGTGATCCGGAGAGGCGCGGTGTCCAGTG 360
QY 361 TACTCGGGGAGGTACAGACAGCCTCAACCTCATTCAGATTAATTCATCCCTTTGAAA 420
    |||||||
Db 361 TACTCGGGGAGGTACAGACAGCCTCAACCTCATTCAGATTAATTCATCCCTTTGAAA 420
QY 421 CTGTGCGCTTCAGAGCCTGAAAGATGAGGGCAGATCTGACAAATTCCTGGCAGTTCCTCT 480
    |||||||
Db 421 CTGTGCGCTTCAGAGCCTGAAAGATGAGGGCAGATCTGACAAATTCCTGGCAGTTCCTCT 480
QY 481 GAGGATGATCCCTGCTTACAGTTCTCCCTGGAAGAAAGCTCAGAAAGAGTGTGAG 540
    |||||||
Db 481 GAGGATGATCCCTGCTTACAGTTCTCCCTGGAAGAAAGCTCAGAAAGAGTGTGAG 540
QY 541 AAAGAAGAAAGAAATGAAAGATTCGCGACAGAGACATCTCTCTTTGCCCAACT 600
    |||||||
Db 541 AAAGAAGAAAGAAATGAAAGATTCGCGACAGAGACATCTCTCTTTGCCCAACT 600
QY 601 TCGTCAAGAAACAAAGCAGAAAGCATACGAGGCGCTCCAGAAAGCTAAGGAAAGTGAAC 660
    |||||||
Db 601 TCGTCAAGAAACAAAGCAGAAAGCATACGAGGCGCTCCAGAAAGCTAAGGAAAGTGAAC 660
QY 661 AACGCTCTCAAGATCTCCCTCTCTGCTGAGCTCCCAAGACAGCAGAGTCCAGCCCTT 720
    |||||||
Db 661 AACGCTCTCAAGATCTCCCTCTCTGCTGAGCTCCCAAGACAGCAGAGTCCAGCCCTT 720
QY 721 CAGAGCAGAGATGAGTGTGCTCTGATGGAAGGCGCTCTTGGCACAGAGCTCTCGA 780
    |||||||
Db 721 CAGAGCAGAGATGAGTGTGCTCTGATGGAAGGCGCTCTTGGCACAGAGCTCTCGA 780
QY 781 CTCTTACACTCAAGATCCGGTCCGGGCTGACCTAGTGAAGTGCCTGTCAAGATGTG 840
    |||||||
Db 781 CTCTTACACTCAAGATCCGGTCCGGGCTGACCTAGTGAAGTGCCTGTCAAGATGTG 840
QY 841 GAGCCCTTGAAGATGTGTGATCACATGCGCAATCATTTTGGGGTGTCTCAAAACAGG 900
    |||||||
Db 841 GAGCCCTTGAAGATGTGTGATCACATGCGCAATCATTTTGGGGTGTCTCAAAACAGG 900
QY 901 ATTCTTTTGGCTTTTGGAGAGTGAACGTCTCTACTGCCAACCCTTAGTACCTTAAG 960
    |||||||
Db 901 ATTCTTTTGGCTTTTGGAGAGTGAACGTCTCTACTGCCAACCCTTAGTACCTTAAG 960
QY 961 CTGAGATGCTGACATCATTTGATGTGTGTCTAGCAAGCTTTTCAGAGGCGCACAG 1020
    |||||||
Db 961 CTGAGATGCTGACATCATTTGATGTGTGTCTAGCAAGCTTTTCAGAGGCGCACAG 1020
QY 1021 ACATCCAGAGAGCTCCGGCTCCGGGTGCAAGGGAAGAAACCAAGATGTTGGAGATC 1080
    |||||||
Db 1021 ACATCCAGAGAGCTCCGGCTCCGGGTGCAAGGGAAGAAACCAAGATGTTGGAGATC 1080
```

```
Db 1021 ACATCCAGAGAGCTCCGGCTCCGGGTGCAAGGGAAGAAACCAAGATGTTGGAGATC 1080
QY 1081 TCATGCTCTCCCTATCTCTCTCTTAAGGTTCATGTCACATATGAGGAAGCATGGGA 1140
    |||||||
Db 1081 TCATGCTCTCTCCCTATCTCTCTCTTAAGGTTCATGTCACATATGAGGAAGCATGGGA 1140
QY 1141 CTCTCTGACACAAAGCTCTCTCTCTTGTGATGGGACAAAGCTTTCAGGCAAGAGCTG 1200
    |||||||
Db 1141 CTCTCTGACACAAAGCTCTCTCTCTTGTGATGGGACAAAGCTTTCAGGCAAGAGCTG 1200
QY 1201 CCAGCTGATCTGGGCTCGAATCCGAGATCTCATGGAAGTGTGGGCTGAAAGCTTCAC 1260
    |||||||
Db 1201 CCAGCTGATCTGGGCTCGAATCCGAGATCTCATGGAAGTGTGGGCTGAAAGCTTCAC 1260
QY 1261 CCGTGTGAGGAGCAAGCCAGACATGAGACATATGCTCCCAATTTTATTTATGTGATT 1320
    |||||||
Db 1261 CCGTGTGAGGAGCAAGCCAGACATGAGACATATGCTCCCAATTTTATTTATGTGATT 1320
QY 1321 TTTTCCGCCCATTAAGGCGCTAACAAGAACTGAATTAAGAACTTGTACTTAATTTCTGG 1380
    |||||||
Db 1321 TTTTCCGCCCATTAAGGCGCTAACAAGAACTGAATTAAGAACTTGTACTTAATTTCTGG 1380
QY 1381 TCGTGGGAGATTGAACCCAGACTATGACATGCTAAGGATGTATGAAGTGGAGGCAAAAC 1440
    |||||||
Db 1381 TCGTGGGAGATTGAACCCAGACTATGACATGCTAAGGATGTATGAAGTGGAGGCAAAAC 1440
QY 1441 CAAGGCAATTAATTAAGCCAGCTAGTGAAGTGTAGTCAAGCAAGTGGCTACTTGG 1500
    |||||||
Db 1441 CAAGGCAATTAATTAAGCCAGCTAGTGAAGTGTAGTCAAGCAAGTGGCTACTTGG 1500
QY 1501 TAGTGTGTGCTGTGTGTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
    |||||||
Db 1501 TAGTGTGTGCTGTGTGTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
QY 1561 GGAACCTTGGCTTCCCTACATTTACATTTACATTTACATTTACATTTACATTTACATTT 1620
    |||||||
Db 1561 GGAACCTTGGCTTCCCTACATTTACATTTACATTTACATTTACATTTACATTTACATTT 1620
QY 1621 CTGTGAAGCTGCTATGTTGAGTGTGTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1680
    |||||||
Db 1621 CTGTGAAGCTGCTATGTTGAGTGTGTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1680
QY 1681 CCGTGGGTTTGTCCGACAGTGTGAGCAGCAGCAGTGTGACTTCTGTGACTGTAGAA 1740
    |||||||
Db 1681 CCGTGGGTTTGTCCGACAGTGTGAGCAGCAGCAGTGTGACTTCTGTGACTGTAGAA 1740
QY 1741 TGCATTTCTGCTGTGAGACACTCGGGTGCAGAAATATTAACAGAGGTGACATGCTG 1800
    |||||||
Db 1741 TGCATTTCTGCTGTGAGACACTCGGGTGCAGAAATATTAACAGAGGTGACATGCTG 1800
QY 1801 AAGCTGAGACTAGTGCAGAAAGTTAAGCAGATTTTCAAGCTTGGGTATCTCTCT 1860
    |||||||
Db 1801 AAGCTGAGACTAGTGCAGAAAGTTAAGCAGATTTTCAAGCTTGGGTATCTCTCT 1860
QY 1861 GCCTGCCAGACTCTAGCCAGTGTGTGTACACACTTCTTGGCATGAGACCTAGTGTGA 1920
    |||||||
Db 1861 GCCTGCCAGACTCTAGCCAGTGTGTGTGTACACACTTCTTGGCATGAGACCTAGTGTGA 1920
QY 1921 CCGGGGCGGATTCGGCGGCGGCTGAG 1946
    |||||||
Db 1921 CCGGGGCGGATTCGGCGGCGGCTGAG 1946
```

RESULT 2

US-09-192-611-1

; Sequence 1, Application US/09192611

; Patent No. 6090561

; GENERAL INFORMATION:

; APPLICANT: Glimmer, Laurie H.

; APPLICANT: Hodges, Martin R.

; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP4S AND METHODS

; NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/192,611
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/755,584
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Giulio A. Decont, Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: HUI-026
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 13..1248
US-09-192-611-1

Query Match 100.0%; Score 1946; DB 3; Length 1946;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1946; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTGTGGAGATGGCGGAACCACTAGAGGAGCTGTCGAGAGTCCCGCGGTGGCGGA 60
DB 1 ACAGTGTGGAGATGGCGGAACCACTAGAGGAGCTGTCGAGAGTCCCGCGGTGGCGGA 60
QY 61 GCGCGTGGAGAGCGCGAGCGCGCGGTGCTGCTGCGCGCGCGCACTTCGCGCT 120
DB 61 GCGCGTGGAGAGCGCGAGCGCGCGGTGCTGCTGCGCGCGCGCACTTCGCGCT 120
QY 121 AGGCTCATTTCCAGACACCGCTGTTGTGAGACTTGGTCACTGACAGCAAGAGGCTTG 180
DB 121 AGGCTCATTTCCAGACACCGCTGTTGTGAGACTTGGTCACTGACAGCAAGAGGCTTG 180
QY 181 GAAGTCCAGACACCACTAGAGAGTCCGCTGCGCGCGCGCTCCCGCGCGGCTAAACCTGAG 240
DB 181 GAAGTCCAGACACCACTAGAGAGTCCGCTGCGCGCGCGCTCCCGCGCGGCTAAACCTGAG 240
QY 241 CAGGACAGCAGACAGTGAAGAGGCGCGAGGCGCGCTGCGGAGCGCGCGCTGACA 300
DB 241 CAGGACAGCAGACAGTGAAGAGGCGCGAGGCGCGCTGCGGAGCGCGCGCTGACA 300
QY 301 TTGGTGGAGAGCGCGCGCGCGCTGCTGAGATCCGAGAGAGCGCGGTGCTCCAGTG 360
DB 301 TTGGTGGAGAGCGCGCGCGCGCTGCTGAGATCCGAGAGAGCGCGGTGCTCCAGTG 360
QY 361 TACTCCGGAGGATACAGAGCAGCTCACTTCACAGTAATTCATCCCTTTAAA 420
DB 361 TACTCCGGAGGATACAGAGCAGCTCACTTCACAGTAATTCATCCCTTTAAA 420
QY 421 CTGTGCGCTTCAAGAGCTGAAGATGAGGAGATCTGACAAATTCGACATTCCTCT 480
DB 421 CTGTGCGCTTCAAGAGCTGAAGATGAGGAGATCTGACAAATTCGACATTCCTCT 480

DB 421 CTGTGCGCTTCAAGAGCTGAAGATGAGGAGATCTGACAAATTCGACATTCCTCT 480
QY 481 GAGATGATGCCCTGCTTCCAGTTCTCCGAGAAAGAGCTCAAGAAAGTGTAG 540
DB 481 GAGATGATGCCCTGCTTCCAGTTCTCCGAGAAAGAGCTCAAGAAAGTGTAG 540
QY 541 AAAGAGAAAGAAATGGAAGATTTCCGAGCAGACATCTCTTCCCGCAACT 600
DB 541 AAAGAGAAAGAAATGGAAGATTTCCGAGCAGACATCTCTTCCCGCAACT 600
QY 601 TCGTCAAGAAACAAAGCAGAAAGCATACGAGCGCTCCGAAAGCTAAGGAAAGTAAAC 660
DB 601 TCGTCAAGAAACAAAGCAGAAAGCATACGAGCGCTCCGAAAGCTAAGGAAAGTAAAC 660
QY 661 AAGGCTCCAGAGATCTCCGCTGCTGAGCGCCCAAGCAGACAGAGTCCAGCCTT 720
DB 661 AAGGCTCCAGAGATCTCCGCTGCTGAGCGCCCAAGCAGACAGAGTCCAGCCTT 720
QY 721 CAGAGCAGAGATGATGAGAGTGTCTAGTGAAGGCGCTGCTTCCACAGAGCTTCGA 780
DB 721 CAGAGCAGAGATGATGAGAGTGTCTAGTGAAGGCGCTGCTTCCACAGAGCTTCGA 780
QY 781 CTCTTTACACTCAAGATCCGCTGCGCGCTGCTAGTGAAGTCTGATGAGATGCG 840
DB 781 CTCTTTACACTCAAGATCCGCTGCGCGCTGCTAGTGAAGTCTGATGAGATGCG 840
QY 841 GAGCGCCCTTCAAGATGAGTGCATGAGCAGATGCGCGCTGCTAGTGAAGTCTGATGAG 900
DB 841 GAGCGCCCTTCAAGATGAGTGCATGAGCAGATGCGCGCTGCTAGTGAAGTCTGATGAG 900
QY 901 ATTCTTTGCTTTTGGAGAGTGAAGTCTGCTTCACTGCGACCCCTAGTACCTTAAAG 960
DB 901 ATTCTTTGCTTTTGGAGAGTGAAGTCTGCTTCACTGCGACCCCTAGTACCTTAAAG 960
QY 961 CTGAGAGTGGCTGACATCATTTGATGTTGTGTGCTAGCAGCTTTCAGAGCCACAGAG 1020
DB 961 CTGAGAGTGGCTGACATCATTTGATGTTGTGTGCTAGCAGCTTTCAGAGCCACAGAG 1020
QY 1021 ACATCCAGAGCTCCGCGCTCCGCGTCCGAGGAGGAAAGCAACCATTTTGGAGATC 1080
DB 1021 ACATCCAGAGCTCCGCGCTCCGCGTCCGAGGAGGAAAGCAACCATTTTGGAGATC 1080
QY 1081 TCACGTCTCTGATGATTCCTCTTAAGTTCATGTCACACTATGAGAGGAGCATGGA 1140
DB 1081 TCACGTCTCTGATGATTCCTCTTAAGTTCATGTCACACTATGAGAGGAGCATGGA 1140
QY 1141 CTCTGTGAGACAAAGCTCTCTCTTCTTGTAGTGGGACAAAGCTTTCAGGCAAGAGCTG 1200
DB 1141 CTCTGTGAGACAAAGCTCTCTCTTCTTGTAGTGGGACAAAGCTTTCAGGCAAGAGCTG 1200
QY 1201 CCAGCTATCTGGGCTGGAATCCGAGATCTCATGCAAGTCTGGGCTGAAGCTTCAC 1260
DB 1201 CCAGCTATCTGGGCTGGAATCCGAGATCTCATGCAAGTCTGGGCTGAAGCTTCAC 1260
QY 1261 CCTGTTGAGAGCAAGCAAGCATGAGACATGAGCATTCCTCAATTTATTTATTTGATTT 1320
DB 1261 CCTGTTGAGAGCAAGCAAGCATGAGACATGAGCATTCCTCAATTTATTTATTTGATTT 1320
QY 1321 TTTCGCCCATTAAGGGCTTAACAGAACTGAATTAAGCTTTATTTATTTCTGG 1380
DB 1321 TTTCGCCCATTAAGGGCTTAACAGAACTGAATTAAGCTTTATTTATTTCTGG 1380
QY 1381 TGCTGGGATTTGAACCCAGACTATGCAATGCTAAGGATGATGAAGTGGAGGCAAAAC 1440
DB 1381 TGCTGGGATTTGAACCCAGACTATGCAATGCTAAGGATGATGAAGTGGAGGCAAAAC 1440
QY 1441 CAAGGCAATTAACCTTTAGCAGAGCTCTAGTACGCTAGTGCAGAGCAAGTGGCTTGG 1500
DB 1441 CAAGGCAATTAACCTTTAGCAGAGCTCTAGTACGCTAGTGCAGAGCAAGTGGCTTGG 1500
QY 1501 TAGTTGTGTGCTCTGTATGTTGTGTGTTGCTGATTTTGGAGCCCTGGGCAATAGAG 1560
DB 1501 TAGTTGTGTGCTCTGTATGTTGTGTGTTGCTGATTTTGGAGCCCTGGGCAATAGAG 1560

LENGTH: 4242 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
US-07-705-490-2

Query Match
Best Local Similarity 59.7%; Score 44; DB 3; Length 4242;
Matches 74; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 243 GGACAGCGAGTGACAGTAAAGGGGGCGGAGGGGCTGCGGAGAGCCCGCGTACATT 302
DB 1642 GGGCGGGCGGGTGACGAGAGCGCGCTGCGAGGGGCGTGGCGACAGCGGCGGCGGCG 1701
QY 303 GGTGCGAGCGGCGGCGGCGGCGGCTGCTGATCCCGAGAGCGCGCGGTGCTCCAGTGT 362
DB 1702 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1761
QY 363 CTCC 366
DB 1762 CACC 1765

RESULT 7
US-07-751-891B-23
Sequence 23, Application US/07751891B
Patent No. 6180337
GENERAL INFORMATION:
APPLICANT: Caskey, C. T.
Nelson, David L.
Pieretti, Maura
Warren, Stephen T.
Oostera, Ben A.
Fu, Ying-hui
TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas D. Paul
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07751.891B
FILING DATE: 29-Aug-1991
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5350
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5325
TELEFAX: 713/651-5246
TELEX: 762829
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 3222 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-07-751-891B-23

Query Match
Best Local Similarity 59.7%; Score 44; DB 4; Length 5222;
Matches 74; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 243 GGACAGCGAGTGACAGTAAAGGGGCGGAGGGGCGTGGCGAGAGCCCGCGTACATT 302
DB 2671 GGGCGGGCGGGTGACGAGAGCGCGCTGCGAGGGGCGTGGCGAGCGGCGGCGGCGGCG 2730
QY 303 GGTGCGAGCGGCGGCGGCGGCGGCTGCTGATCCCGAGAGCGCGCGGTGCTCCAGTGT 362
DB 2731 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2790
QY 363 CTCC 366
DB 2791 CACC 2794

RESULT 8
US-08-455-073A-1
Sequence 1, Application US/08455073A
Patent No. 5876949
GENERAL INFORMATION:
APPLICANT: Gideon Dreyfuss
APPLICANT: Mikiko C. Stomi
APPLICANT: Yan Zhang
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949r1s
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08455,073A
FILING DATE: 31-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: UPN-2201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4362 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..2118
US-08-455-073A-1

Query Match
Best Local Similarity 67.0%; Score 41.6; DB 2; Length 4362;
Matches 59; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 243 GGACAGCGAGTGACAGTAAAGGGGCGGAGGGGCGTGGCGAGAGCCCGCGTACATT 302
DB 12 GGGCGGGCGGGTGACGAGAGCGCGCTGCGAGGGGCGTGGCGAGCGGCGGCGGCGGCGGCG 71


```

1 TITLE OF INVENTION: Sequences and Antibodies Thereto
2 NUMBER OF SEQUENCES: 31
3 CORRESPONDENCE ADDRESS:
4 ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
5 STREET: Two Embarcadero Center, 8th floor
6 CITY: San Francisco
7 STATE: CA
8 COUNTRY: USA
9 ZIP: 94111-3834
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.30
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/724,394A
17 FILING DATE: 01-OCT-1996
18 CLASSIFICATION: 536
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Fitts, Renee A.
21 REGISTRATION NUMBER: 35,136
22 REFERENCE/DOCKET NUMBER: 017957-000100
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: 415-576-0200
25 TELEFAX: 415-576-0300
26 INFORMATION FOR SEQ ID NO: 20:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 246240 base pairs
29 TYPE: nucleic acid
30 STRANDEDNESS: not relevant
31 TOPOLOGY: not relevant
32 MOLECULE TYPE: cDNA
33 FEATURE:
34 NAME/KEY: misc_feature
35 LOCATION: 1..246240
36 OTHER INFORMATION: /note= "HLA-H-CONFIG"
37
38 US-08-724-394A-20

```

This Page Blank (top)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 : Search time 17.87 seconds
(without alignments)
1756.235 Million cell updates/sec

Title: US-09-192-611-2
Perfect score: 2099
Sequence: 1 MAEPLRGPRSGRGARR.....GKELPADLGESGLIEVWG 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154.5	7.4	1217	2 T00270	hypothetical prote
2	145.5	6.9	1072	1 A37221	neurofilament trip
3	140	6.7	1634	2 T26517	hypothetical prote
4	137.5	6.6	854	2 S02003	neurofilament trip
5	134.5	6.4	587	1 WMBETE	65k early nonstruc
6	133	6.3	971	2 T19431	hypothetical prote
7	132.5	6.3	792	2 T29187	hypothetical prote
8	132	6.3	586	1 WMBEDE	65k early nonstruc
9	128.5	6.1	635	2 T09648	nucleolin homolog
10	128.5	6.1	734	2 B42680	nucleolin homolog
11	128	6.1	303	1 Z8BE22	gene 8 protein - p
12	127	6.1	1365	2 S14871	suppressor two of
13	126	6.0	856	2 T16543	hypothetical prote
14	124.5	5.9	1684	2 JMO057	hypothetical prote
15	124.5	5.9	1791	2 T02345	gravin - human
16	122	5.8	1082	2 T15269	hypothetical prote
17	120	5.7	669	2 JC5662	hepatoma-derived g
18	119	5.7	450	2 T08701	hypothetical prote
19	119	5.7	765	2 T49592	neurofilament trip
20	119	5.7	1603	2 S17983	gene posterior sex
21	119	5.7	1611	2 T38236	hypothetical prote
22	118.5	5.6	508	2 T04605	hypothetical prote
23	118	5.6	577	2 T12336	hypothetical prote
24	117.5	5.6	1110	2 I51116	NF-180 - sea lamp
25	117.5	5.6	2187	2 T30826	nascent polypeptid
26	117	5.6	553	2 C75318	hypothetical prote
27	116.5	5.6	714	2 T22454	hypothetical prote
28	116.5	5.6	955	2 S52959	male-specific leth
29	116.5	5.6	974	2 A40580	lodestar maternal-

30	116	5.5	672	2 H86169	hypothetical prote
31	116	5.5	1043	2 A56037	DNA-binding protei
32	115.5	5.5	872	2 S62061	SCD5 protein - yea
33	115.5	5.5	1280	2 T00365	hypothetical prote
34	115	5.5	382	2 T14336	RAD23 protein, iso
35	114	5.4	891	2 G84693	probable prolina-r
36	114	5.4	1274	2 T16251	hypothetical prote
37	113.5	5.4	2361	2 T25752	hypothetical prote
38	113	5.4	611	2 T06458	nucleolin homolog
39	113	5.4	776	2 T20738	hypothetical prote
40	112.5	5.4	95	2 JC4760	SMT3 protein - hum
41	112.5	5.4	675	2 T03744	myod protein inhib
42	112.5	5.4	2109	2 T33247	hypothetical prote
43	112.5	5.4	3942	2 T42730	Bassoon protein -
44	112	5.3	480	3 JC7552	Shb-like adapter p
45	112	5.3	900	2 C96842	hypothetical prote

ALIGNMENTS

RESULT 1
T00270
hypothetical protein KIAA0596 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00270
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A:Reference number: Z14086; MIM:98290545
A:Accession: T00270
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 11217 <NAC>
A:Cross-references: EMBL:AB011168; NID:93043715; PIDN:BAA25522.1; PID:93043716
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0596

Query Match	7.4%	Score 154.5;	DB 2;	Length 1217;
Best Local Similarity	23.5%	Pred. No. 0.086;		
Matches 105;	Conservative 48;	Mismatches 164;	Indels 129;	Gaps 21;
QY 6	RGRGPRSRGGRGARRARGARGRCPRAROSPRLIPDVIVLVSDSEVLEVP	65		
DB 455	RGRGKQGGPSSPQRASG---PNRHQAPSMSPGA---LSDSDKEGDETEELP	506		
QY 66	-----VALRPAKPEQSDSDSEGAEE-----GPAGAPRTLVRRRRRL	106		
DB 507	ALPYLAKSTKKALASVPSPALPRSLSHWMSRAQESGFLDPAPANPGR---RRGRW	563		
QY 107	DPGEPAPVYVSGKVOSSLNIPDNSSLKICPS--EP-EDPADLITNSGSSPDALPS	163		
DB 564	QPG-----VELSVRSMDLRLQLETLAPSLQDPQSDSLAIFSGRRKMQEALFT	612		
QY 164	G-----SPWRKKLRKCC-----EKPE-----KMEPEPPDDISPLPOPS---SRNK	201		
DB 613	SLTSQNKPPPPQASQPCSTPHIIRLLSQEBGVFAQULEPAPIDGIVYEPSPNPMDT	672		
QY 202	SRKHTKALQK-----LREVNKRLOD-----LRCSLSPKQKOSPALQSTIDDEVLT	245		
DB 673	SEFQVQAPARAGTLGRVYVPGSRSSSEKHPDSCVSDVSSCLSPSEH--PTEDSESTPLS	730		
QY 246	VEGPVLPOSSRLFTLKRCRADVRLPYRMSEPLQNVVDMAHNLGVSPNRILLFESE	305		
DB 731	VDG-----ISSDLEEPGEDEEEEGGNGP-----YGLQ	762		
QY 306	LSPTAP-----STLKIGVADIIIDCVLASSEATETSOELRLRYQKREKHMLE	355		
DB 763	GSP-QITDQEQFLKQHEFTLASCAAPGAP-VQYPERSESISIRFLQVOTRLREP--	818		

Db 409 SDDKKKVVAKPTSAKATPKPAKKRADSSSDDEAPAKKTPAKAAPKASKAE---SS 465
 QY 226 LSPKQHOSPALQSTDDVEY 243
 Db 466 SDSSDDEKPAKSTPAKI 483

RESULT 7

T29187

hypothetical protein C55C3.3 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T29187

R/Messner, J.; Steiljes, L.

Submitted to the EMBL Data Library, April 1996

A/Description: The sequence of *C. elegans* cosmid C55C3.

A/Reference number: Z20585

A/Accession: T29187

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-792 <MOE>

A/Cross-references: EMBL:U53335; PIDN:AAA96170.1; GSPDB:GN00022; CESP:C55C3.3

A/Experimental source: strain Bristol N2; clone C55C3

C/Genetics:

A/Gene: CESP:C55C3.3

A/Map position: 4

A/Introns: 17/1; 46/3; 139/2; 173/1; 204/3; 250/3; 266/3; 286/2; 327/1; 362/3; 393/3; 43

Query Match 6.3%; Score 132.5; DB 2; Length 792;
 Best Local Similarity 20.9%; Pred. No. 1.1;
 Matches 90; Conservative 50; Mismatches 135; Indels 155; Gaps 19;

QY 11 RSRGGRGARRAGRGRCPRARQSPARLIPVTVLVDLSDEDEVLEVPVVARLP 70
 Db 309 RRRGNGRIKKKKPAK---PLKRITP-----NNSDEKLKKMKSP----- 343
 QY 71 APAKPEODSDSD-----SEGAEGPAGAPRTLVRRRRR-----LLDPG 109
 Db 344 -PKKPLQATOSDILQPLPSNNAMTVPESSKESANNSPOSSKINLRNGGRVYNTDLDCS 402
 QY 110 EAPVVPYVSGVQSSSLNIPNSSLKLCSEPEDEADLTNGSSPSEDALPGSGPWK 169
 Db 403 EARAIP-----KC-DKLTDFQTSWSEQSPA--DAISAVAP-TK 437
 QY 170 KLRKCKEKKMEFPD-----ODI-----SPLPQSSRNK----- 201
 Db 438 LTGQSQSSEKQEIEMKMSFRELQDSKFKPAEAFIPKPEERKSGSTQSTTEPTKLQSG 497
 QY 202 SRKHTALQKLRVYKRIQD--LRSLSPKQHOSPALQSTDD----- 241
 Db 498 SPSEKQEIQMPNFSFRELQDSKFKPAPSEKQPAQSELPSPVPGNNRRNSETSOVD 557
 QY 242 -----EVVLVEGPVLPQSSRLFTLKIRCADLVRLPKRKSEPLQVNVDH-----MAN 288
 Db 558 TISVPPTKLVTGQSPSEKQEIQMPNFSFRELQDSKFKPAPKPKPVERGSAEQSMSE 617
 QY 289 HLGVPKRIILLFGESELSPTAPSTLK-----LGVADIIDCVLA-----SS 331
 Db 618 PLGIS-----RYAEG-SPIAKRPRSPQLQAPLETALPTTIDAPTAIETAIENSAEFSS 672
 QY 332 SEATETSOEL 341
 Db 673 SHSMDPNSNL 682

RESULT 8

WMBEDE

65k early nonstructural protein - human cytomegalovirus (strain AD169)

N/Alternate names: UL84 protein

C/Species: human cytomegalovirus, human herpesvirus 5

A/Note: host Homo sapiens (man)

C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
 C/Accession: S09848
 C/Chee, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horsnell, T
 M.; Barrell, B.G.
 Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
 A/Title: Analysis of the protein-coding content of the sequence of human cytomegalovi
 A/Reference number: S09749; MUID:90269039
 A/Accession: S09848
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-586 <CHE>
 A/Cross-references: EMBL:X17403; NID:959591; PIDN:CAA35358.1; PID:91780862
 A/Note: this sequence was submitted to the EMBL Data Library, December 1989
 C/Superfamily: human cytomegalovirus 65k early nonstructural protein
 C/Keywords: leucine zipper; nonstructural protein
 F:114-135/Region: leucine zipper motif
 F:324-372/Region: leucine zipper motif

Query Match 6.3%; Score 132; DB 1; Length 586;
 Best Local Similarity 23.6%; Pred. No. 0.82;
 Matches 104; Conservative 62; Mismatches 145; Indels 130; Gaps 25;

QY 6 RGRGPRSR---GGRGARRAGRGRCPRARQSPARLIPVTVLVDL--VSDSEYV 55
 Db 11 RARRPRARRGGGGGGSNSRHS-GKCRORRALSLAPLFTLTATTTTMMGVASTDDDS 69
 QY 56 LEVADPVEVVARIPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLLDGEAVVP 115
 Db 70 LLKTPDEL-----DKYSGSPQITL-----TLDKHDIRQPR 101
 QY 116 VYSRK---VQSSNLNLPDN-----SSLLKLCSEPEDEADLTNGSSPSEDALPS 163
 Db 102 VHRGTYHLQLHDLRPEELRDPQILSLTPQLQ--GEANDESG--TAPATLDEETAS 157
 QY 164 GSPWRKRLKCKEKKMEFPDQDISPLPQSSRNKSKKHTDALQKREYKRIQDLR 223
 Db 158 HEPEKKK---EKQEKKEE---DED-----DRNDDKER---GLICVSNESDVR 197
 QY 224 SCLSS--PKQHOSPALQSTDEDEVLPVLEGPVLPQSSRLFTLKIRCADLVRLPKRKSEPLQVNV 281
 Db 198 PAFSLFPARPGCHILKRVYDQ-QLTRMAIVRLSLNLFALRI-ITPLKRLPLR-----RK 250
 QY 282 VYDHMANHLGVSPNRILLFGESELSPTAPSTLKLGADIIDCVLASSSEATETSOEL 341
 Db 251 AAHTTALH-----DCLALHLPFLTFEPTLDINNTEAASVAD---TAESTDA-DLTPTL 301
 QY 342 RLRYQCKEKQMLETISPSPLKVLMSHYEAMGLSG-----HKLS----- 383
 Db 302 TVRVR-----HALCWHVEG-GISGPRGLTSRISARLSETTAKTLGP 342
 QY 384 FFPDGTLSGKELPADLGLES 404
 Db 343 SVFGRELDPNESPDLTLSS 363

RESULT 9

T09648

nucleolin homolog num1 - alfalfa

N/Alternate names: num1 protein

C/Species: *Medicago sativa* (alfalfa)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C/Accession: T09648

R/Boegre, L.; Jonak, C.; Mink, M.; Meskine, I.; Traas, J.; Ha, D.T.C.; Swoboda, I.;

Plant Cell 8, 417-428, 1996

A/Title: Developmental and cell cycle regulation of Alfalfa nucleolin a plant homolog of

A/Reference number: Z16796; MUID:9631876

A/Accession: T09648

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-635 <BOE>

A/Cross-references: EMBL:X08845; NID:q1279562; PIDN:CAA61298.1; PID:q1279563

C/Genetics:


```

Db      173  AAELI-----NIPDYQKEEDAQMGLVPPAV-----GAD 200
Oy      268  LVRLVVRKSEPLQNVVDHMANHLVSPNR---ILLFGESLSPTATSTKLGVADIIID 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      201  IMRLPEKSSAALM-----YHIGANPEKARQLLMDGOSAL----- 235
Oy      325  CVVLASSPATETSOELRLVQGREKHOMLEISLSP--DSP 364
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      236  -----IELRLSERLTLKPRK-----QISSAPPADQPI 264

RESULT 12
SI4871
Suppressor two of zeste protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 06-Jan-1995 #sequenceRevision 06-Jan-1995 #textChange 21-Jul-2000
C:Accession: SI4871; SI6845
R:Brunk, B.P.; Adler, P.N.
submitted to the EMBL Data Library, November 1990
A:Description: The Drosophila regulatory gene suppressor two of zeste encodes a large z
A:Reference number: SI4871
A:Accession: SI4871
A:Molecule type: mRNA
A:Residues: 1-1365 <BRU>
A:Cross-references: EMBL:X56799; NID:g8528; PID:g8529
R:Brunk, B.P.; Adler, P.N.
Nucleic Acids Res. 19, 3149, 1991
A:Title: The sequence of the Drosophila regulatory gene Suppressor two of zeste.
A:Reference number: SI6845; MUID:91279476
A:Accession: SI6845
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-612,614-784,'N',786-830,'R',832-965,967-1064,'E',1066-1096,'N',1097-1286,'
A:Cross-references: EMBL:X56798; NID:g8526; PID:CAA40134.1; PID:g8527
C:Keywords: DNA binding; nucleus; zinc finger
C:Genetics:
A:Gene: FlyBase:Su(2)2
A:Cross-references: FlyBase:FBgn0008654
A:Introns: 55/1; 139/2; 219/3
C:Superfamily: Drosophila suppressor protein of zeste; RING finger homology
C:Keywords: DNA binding; nucleus; zinc finger
F:31-79/Domain: RING finger homology <RNG>

Query Match      6.1%; Score 127; DB 2; Length 1365;
Best Local Similarity 23.1%; Pred. No. 4.6;
Matches 84; Conservative 47; Mismatches 125; Indels 108; Gaps 16;

Oy      110  EAPVVPVYSGKQSSINLIPDNSSILKLCPS-EPDEADLTNIGSSP--SEP--DALPSG 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      220  EEPMAFCYRIILLYDDEQTKNDENNLSRINODIEPHSVRSKASAVFAEDLESEIDSG 279
Oy      165  SPWRKRLKRCCKEKEKMEFPDODISPLPQSSRNK---SRKTHEA-----LQKLE 214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      260  SP-RKSVCK-----TPKVSRSKRNKRLTSSKRAEPSPVSNKSLRS 323
Oy      215  VNRKLDL-----RSCLSPKQHQSPALQSTDEEVLVEGVPVLPQS---SRLEFTL 260
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      324  NDMRYSDYAVSKVSKSEPEQOFLPREREQPLFANTNIVASIPSQLRKSVDADEPDL 383
Oy      261  KIRCADVLRLP-----VRMSEPL-----QNVVDHMANHLGV 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      384  KYANKRGVGHLEPKLKIELNSMKSLMPLSAGRLDETSCESSCSAODLDLETYAKNIGL 443
Oy      293  SPNRILLFLFG-----SESLSPATP-----STKLGVADIIDCVVLASSPATETS 338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      444  KPIEOPPLQOSASNPQSKSPNAPSMSGSSSTNGSSSSISLTAD-----ASTSTSSS 496
Oy      339  QELRLR-----VQGREKHOMLEISLSPDPLVYLMSHYEAMGLSGHKLSPFPGTK 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      497  HRRKKKSKSKEPKDANGKRRKKLAHLSISQTDOKMKVKIT-----AKRNHKLDFKRSHSL 550
Oy      391  LSGK 394

```

```

Db      551 ASGE 554

RESULT 13
716543
hypothetical protein K03C7.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C:Accession: J16543
R:Leimbach, D.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid K03C7.
A:Reference number: Z18532
A:Accession: J16543
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-856 <LEI>
A:Cross-references: EMBL:U40059; NID:g1055170; PID:g1055171; PIDN:AA0138.1; CESP:K03C7.1
A:Gene: CESP:K03C7.1
A:Introns: 176/1; 632/2; 653/1; 717/3; 746/3; 791/3
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match      6.0%; Score 126; DB 2; Length 856;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 85; Conservative 55; Mismatches 153; Indels 104; Gaps 17;

QY  2 AEPILRGPRSRSGCGARRGARNGRCPRARQSPARLIPDIVIVDVSDEEVLVADP 61
Db  362 ADPL--GGPKKKDKPLAKKAPAKK--PTDKPKK-----DAPKDAKPTPEEPKAP 427
QY  62 V-----EVPVRLPAPAKPCODSDSEGAEGPAGA-----PRTIVRR 100
Db  428 VAPKKKKRPWEEDPDPEPEPEADFTMPAKKKPTDTPRAD-----PLGGPNKKDKRLKAKKA 482
QY  101 RRRRLD---PGEADVVPVYSGKVQSSLNLIIPDNSS-----LLKICPSEPEDEADLTNSGS 153
Db  483 PTKKADADPKPSEEPKGVAPAKKMPWEEDPDDEADFTVPIKPGCEDEDEPADDEE 542
QY  154 SPSEDDALPSSGPMKKLTKKCKEKKKEEPPODDISPLQP----- 196
Db  543 PEED--PADEEPKKKPKKHKRRKKPKVVEPEKEPTPEVPVPAKPKIADIKAPEE 599
QY  197 -----SSRKSKRKTAEALQKLEVENKRLDQLRSCSPKQHOHSPAL----- 236
Db  600 PIPMPKPKETIAERKKKEERIPALRYAKK-PRELEVYIPVYIPME-QTAAALITDEGMCAF 657
QY  237 ---QSTDEVVVLVEGVPVLP---QSSRLFTL-----KINCRADLVRLPYRMSEPLQNVVDH 285
Db  658 GKSRANAEVNFNGDKPIVOGAVDSKTVIPLWNDESKCANRSGMTAFGA-PRRIDQNVVDH 716
QY  286 MANHILGVSPNRIILLFGESELSPTATPSTLKGAVDI 322
Db  717 --HKKSQGIIPLLAKGTYNPHGEYGTIRROTADV 750

RESULT 14
JM0057
gravin - human
C:Species: Homo sapiens (man)
C:Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C:Accession: JM0057
R:Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A>Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelia
A:Reference number: JM0057; MUID:98269042
A:Accession: JM0057
A:Molecule type: mRNA
A:Residues: 1-1684 <SAT>
A:Cross-references: DDBJ:AB003476; NID:g2081606; PIDN:BA019927.1; PID:d1020716; PID:g2081606; This protein regulates cell growth.

```

F:433-439/Region: nuclear location signal
 F:522-527/Region: nuclear location signal
 F:591-596/Region: nuclear location signal
 F:671-676/Region: nuclear location signal

Query Match 5.98; Score 124.5; DB 2; Length 1684;
 Best Local Similarity 20.7%; Pred. No. 8.5;
 Matches 94; Conservative 70; Mismatches 171; Indels 119; Gaps 21;

QY 43 VLVVDVSDEDEVLEVPVEVVARLPAPAKPEQSDSD----- 82
 DB 32 VVHDTDGGQETPEITIQIPSSSESNLEELTPTESQANDIGKKVKEFKVGFKTVPKDK 91
 QY 83 -----SEGAAGPAGA-----PRTLVRRRRLLDPGAPVVPVYSGK-VQSS 124
 DB 92 TEKPDVQLLVKKNDEGGAAGAGDHKPSLGAAGAAKSEPKOSTEKPEITLKREQSH 151
 QY 125 LNLIP--DNSSLKLCSPSEPD--EADLTNCGSPSEDDALPSGSP-----WRK 169
 DB 152 AEISPPASGQAVECEKEGEKEKPSKSAESPTSPVTSSTGTFKKFTQGWAGWRK 211
 QY 170 K--LRKKEKE-----EKMEEPDODISLPQPSRNRKSKHTEALQKREYNKRLODR 223
 DB 212 KTSFKPEDEVEASEKKKEQEPK----VTEEDGKAEEASEKLTASEQAHPQ--EPAE 265
 QY 224 SCLSPK--OHOSPALOSTDEEVLVVEGP-----VLPOSSRLFTLKIRCADLVRLPVRMS 276
 DB 266 SAHEPRLSAETKEVELPS-EROVSGSGSPSEKPAPLATEVFDEKIEVHOEEVAEVAHV 324
 QY 277 EPLQNVVDHMANHLGVSPNRILLFGESELSPTAPSTLKLGVADIIDCVVIASSSEATE 336
 DB 325 TVEERTEEQ-----KTEVEETA-----GSVAEELVEMDAEPQEA 360
 QY 337 TSQEL-RLR--VQKEKQHMLEISLSPDPL-----KVLMHYEAPAGLSGHL 382
 DB 361 PAKELVTLKETVCVSGEDPTQAGD--LSPDEKVLKPRGVSEVEMLSOERMKVQSGPL 418
 QY 383 SFFPDGT---KLSGKELPADLG---LESGDLIEV 410
 DB 419 KKLFTSTGLKTLGSKKQKGRGGDESGEHTQY 452

RESULT 15

T02345
 hypothetical protein KIA0324 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
 C:Accession: T02345
 R:Ricke, D.O.; Bruce, D.; Munde, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
 re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
 Submitted to the EMBL Data Library, March 1998
 A:Description: Sequencing of human chromosome 16p13.3.
 A:Reference number: Z14564
 A:Accession: T02345
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1791 <RIC>
 A:Cross-references: EMBL:AC004493; NID:g2996648; PTDN:AC08453.1; PTD:g2996650
 C:Genetics:
 A:Map position: 16
 A:introns: 1610/2; 1706/2
 A:Note: KIA0324

Query Match 5.98; Score 124.5; DB 2; Length 1791;
 Best Local Similarity 25.3%; Pred. No. 9.2;
 Matches 72; Conservative 26; Mismatches 96; Indels 91; Gaps 13;

QY 34 SPARLIPPTVAVLVSDDEVLEVPVEVVARLPAPAKPEQSDSDSEGAAGPAGA 93
 DB 414 SSELSPDAVEKAGKSSNQ-----SISFVLDAVPTPTSRER-----SSASSPEMKDGL 463

QY 94 PRTLVRRRRLLDPGAPVVPVYSGKVQSSLNLIPDNSSL-----LKLCPSEPD 144
 DB 464 PRTPSRRSR-----SGSSPGLRDGSG-----TPSRHSLSGSSPGMKDIPRTPSRGRS 510
 QY 145 EADLTNCGSPSEDDALP-----SGSPWRKKLRKKC-----EKKEKMEEPDQ----- 188
 DB 511 ECD-----SSP-EPKALPOTPRPSRSPSPPELNKCLTPQRENSGSSSVQKTVARTP 564
 QY 189 -----DISPLPQSSR-----NKSRKHTEALQKREYNKRLODRSCLS 227
 DB 565 LQGRSRSSSQEELDVKPSASFOERSESDSPDSKAKTRTPLRQRRSGSSPEVDSKRLS 624
 QY 228 PKHQSS-----PALQSTDEEVLVVEGP---VLPOSSR 256
 DB 625 PRSRSGSSPEYKDKPRAAPRAQSGSDSSPEPKAPAPRALPRSR 669

Search completed: September 11, 2001, 08:49:55
 Job time: 89 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:49:31 ; Search time 10.66 Seconds

(without alignments)
1323.946 million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099

Sequence: 1 MAEPLRGPRGSRGGRARR.....GKELPADGLGSGDLIEWWG 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	135.5	6.5	831	1 NPH_RAT	P16884 ratius norv
2	134.5	6.4	587	1 UL84_HCMVT	P29839 human cytom
3	132	6.3	586	1 UL84_HCMVA	P16737 human cytom
4	128.5	6.1	704	1 NP14_RAT	P41777 ratius norv
5	128	6.1	303	1 VG08_BPP22	P26748 bacterioph
6	127	6.1	1365	1 SUZ2_DROME	P25172 drosophila
7	122.5	5.8	1781	1 AKAC_HUMAN	O02952 homo sapien
8	121	5.8	1109	1 TCF8_RAT	O02947 ratius norv
9	119	5.7	1603	1 PSC_DROME	P35820 drosophila
10	116.5	5.6	1039	1 MS11_DROME	P50535 drosophila
11	116.5	5.6	1061	1 LDS_DROME	P34739 drosophila
12	116	5.5	1043	1 TCF8_MESAU	O60542 mesocricetu
13	115.5	5.5	872	1 SCDS_YEAST	P34728 saccharomyc
14	114.5	5.5	1411	1 TCOF_HUMAN	O13428 homo sapien
15	113.5	5.4	1280	1 DYNM_RAT	P28023 ratius norv
16	112.5	5.4	95	1 SM32_HUMAN	P18858 homo sapien
17	112	5.3	919	1 DNLI_HUMAN	P58555 homo sapien
18	112	5.3	943	1 CENC_HUMAN	O03188 homo sapien
19	112	5.3	1004	1 PTPY_RAT	O63475 ratius norv
20	111.5	5.3	1170	1 XPG_MOUSE	P35669 mus musculu
21	111	5.3	741	1 BSG2_DROME	P13027 zea mays (m
22	111	5.3	897	1 EP15_MOUSE	P11929 drosophila
23	111	5.3	1281	1 DYNM_MOUSE	P42557 mus musculu
24	110.5	5.3	919	1 SYN2_HUMAN	O08788 mus musculu
25	110.5	5.3	2517	1 NCR2_HUMAN	O9Y618 h nuclea
26	110	5.2	542	1 TULI_HUMAN	O00294 homo sapien
27	110	5.2	612	1 ARRS_MAIZE	P13027 zea mays (m
28	110	5.2	883	1 E7AB_DROME	P11536 drosophila
29	110	5.2	1206	1 FM14_MOUSE	O05859 mus musculu
30	110	5.2	2150	1 SDG3_CAEEL	P34706 caenorhabdi
31	109.5	5.2	536	1 GAG_MIVCB	P27460 cas-br-e mu
32	109	5.2	974	1 YMB4_CAEEL	O03601 caenorhabdi
33	109	5.2	1089	1 Y553_HUMAN	O9ukj3 homo sapien

34	108.5	5.2	487	1 ATF2_CHICK	O93602 gallus gall
35	108.5	5.2	505	1 TUB_MOUSE	P50566 mus musculu
36	108.5	5.2	1117	1 TCF8_MOUSE	O64318 mus musculu
37	108	5.1	487	1 ATF2_HUMAN	P15336 homo sapien
38	108	5.1	487	1 ATF2_MOUSE	P16951 mus musculu
39	108	5.1	673	1 FXR2_HUMAN	P51116 homo sapien
40	108	5.1	1402	1 IF4G_RABIT	P41110 oryctolagus
41	107.5	5.1	532	1 ICP0_HYVEB	P28990 equine herp
42	107	5.1	487	1 ATF2_RAT	O00969 ratius norv
43	107	5.1	650	1 NUC1_XENTIA	P20397 xenopus lae
44	107	5.1	1033	1 DYNM_CHICK	P35458 gallus gall
45	106.5	5.1	610	1 ARLC_MAIZE	P13526 zea mays (m

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	831 AA.
RESULT	1			
NPH_RAT				
ID	NPH_RAT			
AC	P16884: 063368;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)			
DE	(NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).			
GN	NEFH OR NPH.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=89065087; PubMed=3143606;			
RA	Breen K.C., Robinson P.A., Wion D., Anderson B.H.;			
RT	"Partial sequence of the rat heavy neurofilament polypeptide (NF-H).			
RT	Identification of putative phosphorylation sites.";			
RL	FEBS Lett. 241:213-218(1988).			
RN	[2]			
RP	SEQUENCE OF 37-831 FROM N.A.			
RX	MEDLINE=88309090; PubMed=2457365;			
RA	Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,			
RA	Jolles P.;			
RT	"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and			
RT	in situ detection.";			
RL	Biochem. Biophys. Res. Commun. 154:1099-1106(1988).			
RN	[3]			
RP	SEQUENCE OF 1-89 AND 243-313 FROM N.A.			
RX	MEDLINE=87080760; PubMed=2878828;			
RA	Robinson P.A., Wion D., Anderson B.H.;			
RT	"Isolation of a cDNA for the rat heavy neurofilament polypeptide			
RT	(NF-H).";			
RL	FEBS Lett. 209:203-205(1986).			
RN	[4]			
RP	SEQUENCE OF 318-831 FROM N.A.			
RX	MEDLINE=89184647; PubMed=2928342;			
RA	Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,			
RA	Smolowitz M., Carroll Z., Emanuel B.S., Breiter J., Rubin L.;			
RT	"Cloning of a cDNA encoding the rat high molecular weight			
RT	neurofilament peptide (NF-H): developmental and tissue expression in			
RT	the rat, and mapping of its human homologue to chromosomes 1 and			
RT	22.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).			
CC	-I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,			
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.			
CC	NEFH HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT			
CC	SUBSERVED BY THE TWO SMALLER NF PROTEINS.			
CC	-I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P. NPH IS			
CC	PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS			
CC	THOUGHT THAT PHOSPHORYLATION OF NPH RESULTS IN THE FORMATION OF			
CC	INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE			
CC	OF AXONAL CALIBER.			

```

Db      646 AKPE-----PKKV-----EEEKTP-ATPKE-----VKEKNDEA 674
QY      335 TETSOELRLRVQGEKQHOMLEISLSPDSLKLVLMSHVEANGLSCHKISFFPDGTSLGSK 394
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      675 PKAQO-----KKPAKEKEPLTE--KPKDSPCAKKEAKERKAAPAE-----E 715
QY      395 ELPADLGIE 403
       | : : : : :
Db      716 ETPAKLGVK 724

RESULT 2
ID     UL84_HCMVT          STANDARD:          PRT:          587 AA.
AC     P29839;
DT     01-APR-1993 (Rel. 25, Created)
DI     01-APR-1993 (Rel. 25, Last sequence update)
DE     01-APR-1993 (Rel. 25, Last annotation update)
DE     65 KDA EARLY NONSTRUCTURAL PROTEIN (UL84 PROTEIN).
GN     UL84.
OS     Human cytomegalovirus (strain Towne).
OC     Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC     Betaherpesvirinae; Cytomegaloviruses.
OX     NCBI_Taxid=10363;
RN     [1]
RP     SEQUENCE FROM N.A.
RX     MEDLINE=92114132; PubMed=1309892;
RA     He Y.S., Xu L., Huang E.S.;
RT     "Characterization of human cytomegalovirus UL84 early gene and
       J. Virol. 66:1098-1108(1992)."
CC     CC
CC     This SWISS-PROT entry is copyright. It is produced through a collaboration
CC     between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC     the European Bioinformatics Institute. There are no restrictions on its
CC     use by non-profit institutions as long as its content is in no way
CC     modified and this statement is not removed. Usage by and for commercial
CC     entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC     or send an email to license@isb-sib.ch).
CC     -----
DR     EMBL: M81432; AAAA5947.1; -.
DR     PIR: A41608; WMBETE.
KW     Nonstructural protein.
FT     DOMAIN              9         ARG-RICH (BASIC).
FT     DOMAIN              162        LYS-RICH (BASIC).
FT     DOMAIN              171        ASP/GLU-RICH (ACIDIC).
FT     DOMAIN              171        183
SQ     SEQUENCE   587 AA; 6538 MM; 13C170EA41FB220B CRC64;

Query Match           6.4%; Score 134.5; DB 1; Length 587;
Best Local Similarity 24.4%; Pred. No. 0.29;
Matches 105; Conservative 58; Mismatches 159; Indels 109; Gaps 22;

QY      6 RGGRPRSR----GGGARARNGARGCRPARQS-----PALIPDVIVLDL--VSDSDEV 55
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      11 RARRPRARRGGGGVGSSNSRRHS-GCKRRORRALSAAPLTFLATTTTTMMGVASTDDDS 69
QY      56 LEVADPVENVYPARLPARPARKPPQDSDSDEGAAGPAGAPRTLVRRRRRRLRLDPGAPVPV 115
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      70 LLKLTPTDEL-----DKHSGSPOTIL-----TLTDKHDIROPR 101
QY      116 VYSGK---VOSSLNLIPDN-----SLLKLCPSPEDEADLTNNGSSPEDDALPS 163
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      102 VHRSTYLIQHLMDLRBELNDPFOILLSTLOLGEANGESQ---TAAATQSEETA-- 156
QY      164 GSPMRKLLIKCEKEEKMEEFPPODISLPDPSSRNKSRRKHTALQLKREYNKRLLDLR 223
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      157 ----SHELEKKEKEKEKEE--DED-----DRNDRER-----GLTCVSNEDSDVR 198
QY      224 SCLS--FKQHSPLAQSTDEEVVLVEGVLPQSSRLFLTKRCRADYLRLVVRKSEPLQN 281
Db      199 PAFSLFAPRPCCHILRSVID-QLTRMAIVLSLNLFRLI-TTRPLRKVPLRKAAHNT 256

```


FT REPEAT 479 490 ACIDIC SERINE CLUSTER 9.
 FT REPEAT 524 535 ACIDIC SERINE CLUSTER 10.
 FT REPEAT 559 570 ACIDIC SERINE CLUSTER 11.
 FT MOD_RES 567 567 PHOSPHORYLATION (BY CK2).
 FT VARIANT 150 150 MISSING (IN NOBPL40B).
 SQ SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;

Query Match 6.1%; Score 128.5; DB 1; Length 704;
 Best Local Similarity 23.2%; Pred. No. 0.83;
 Matches 71; Conservative 35; Mismatches 105; Indels 95; Gaps 12;

QY 33 QSPARLIPPLVLDVLSDESE-----VLEVADPEVPPVRL--PAPAKPEDSDSDSE 84
 DB 317 QSPKKAATOTOPADSSADSSSESEKKTPTAKTVSKTPAKAPAKKKAESSSDS 376
 QY 85 GAAGAGAPRTLLVRRRRRLDPGEAPVYV-----YSGKVQSSLNL 127
 DB 377 DSDSDEAPAKPVSKATSPKSPATVPKPPAKAVATPKPAGSGQKPSRKADSSSE 436
 QY 128 IPDNSS-----LLKICPSPDEEA-----DLTNGSSPSSED----- 159
 DB 437 EESSSSSEETKTSVTPPKARVYAKAPSLPAKQAPRAGDSSSDSSSESEKKTTPK 496
 QY 160 -----ALPSGSPMKRLRK-----KCEKEKKMEPPDODISPLQ----- 195
 DB 497 PPAKKAAGAAYPPKPPVKAALAESSSSSSSDSSSESEKKT-----PKSKATPKPQAGKA 552
 QY 196 ---PSSRN-KSRKHTPALOKLREVNKRLDLSCLSPKQHSPLASTDDEVVLVEGPVL 251
 DB 553 NGVPASONGAGKESEEEEDTEONKKAAGTKPGSGKKRKN-----ETADEA-----AT 602
 QY 252 POSSRL 257
 DB 603 POSKKV 608

RESULT 5

VG08_BP22
 ID VG08_BP22 STANDARD; PRT; 303 AA.
 AC P26748;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SCAFFOLDING PROTEIN (PROTEIN GP8).
 GN 8.
 OS Bacteriophage P22.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
 OX NCBI_TaxID=10754;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
 RX MEDLINE-91306435; PubMed-1853558;
 RA Eppeler K., Wyckoff E., Coates J., Parr R., Casjens S.;
 RT "Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging";
 RT Virology 183:519-538(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kropinski A.M.B., VanderByl C.S.;
 RT "The completed sequence of genome of Salmonella phage P22";
 RT Submitted (Dec-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR SUCCESSFUL CONDENSATION OF DNA WITHIN THE
 CAPSID. THE INTERIOR OF THE PROHEAD IS FILLED WITH THE GP8
 PROTEIN. THE SCAFFOLDING PROTEIN IS LOST FROM THE STRUCTURE
 DURING PACKAGING.
 CC -1- INDUCTION: THE SCAFFOLDING PROTEIN NEGATIVELY REGULATES ITS OWN
 SYNTHESIS WHEN IT IS NOT ASSEMBLED INTO PROHEADS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M59749; AAA72962.1;
 DR EMBL; AF217253; AAF75046.1;
 DR PIR; D40474; Z8BP22.
 KW Late protein.

SQ SEQUENCE 303 AA; 33564 MW; B86BD625D5011ABE CRC64;

Query Match 6.1%; Score 128; DB 1; Length 303;
 Best Local Similarity 21.3%; Pred. No. 0.32;
 Matches 73; Conservative 49; Mismatches 100; Indels 120; Gaps 15;

QY 41 DTVLVLVSDS--DEEVLEVADPEVPPVRLPAPAKPEDSDSDSGAAGPAGARTLV 98
 DB 25 DSLVYNADNMGAGEGFETV-----LKDETAPKODPKKMEFAARRKE 69
 QY 99 RRRRRRLDPGEAPVYVYSGKVQSSLNLIPDNSSLLKICPSEPEDEDLITNGSSPSSED 158
 DB 70 RKQRELEQMEAE-----VKRGELPESLRYNP-----LPQPDINATLSEGLAKYDY 118
 QY 159 D---AL---PSGSPMKRLK---KCEKEKKMEPPDODISPLPSSRNKSKKTE 207
 DB 119 DNSRALAEFAANTEMLMKAQDARSNAVAEOGRKTOEFTQOSAOYV-----EAARKHYD 172
 QY 208 ALQKLEVNKRDLDSLSPKQHSPLASTDDEVVLVEGPVLPOSSRLFTLKICRAD 267
 DB 173 AAEKL-----NIPDYQEKEDAPMLPPAV-----GAD 200
 QY 268 LVRLPYRMSEPLQNVVDHMANHLGVSPNR---LTLFGSELSPTATPSTLKGVAIDI 324
 DB 201 IMRLPEKSAALM-----YHLGANPEKARQLANDGOSAL----- 235
 QY 325 CTVLASSSEATETSOELRLRVGCKEKQMEISLSP--DSPL 364
 DB 236 -----IELTRLSERLTLKPRGK-----QISSAPPADQPI 264

RESULT 6

SU22_DROME
 ID SU22_DROME STANDARD; PRT; 1365 AA.
 AC P25172;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SUPPRESSOR 2 OF ZESTE PROTEIN (PROTEIN POSTERIOR SEX COMBS).
 GN SU(2)2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTON-S;
 RX MEDLINE-91279476; PubMed-2057369;
 RA Brunk B.P., Adler P.N.;
 RT "The sequence of the Drosophila regulatory gene Suppressor two of
 RT zeste";
 RT Nucleic Acids Res. 19:3149-3149(1991).
 CC -1- FUNCTION: REGULATES EXPRESSION OF THE HOMEOTIC SELECTOR GENES BY
 CC INFLUENCING HIGHER-ORDER CHROMATIN STRUCTURE THROUGH INTERACTION
 CC WITH OTHER PROTEINS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

Query Match	6.1%;	Score 127;	DB 1;	Length 1365;
Best Local Similarity	23.1%;	Pred. No. 2.3;		
Matches 84;	Conservative 47;	Mismatches 125;	Indels 108;	Gaps 16;

RESULT	7
AKAC_HUMAN	
ID	AKAC_HUMAN
AC	002952; Q99970; Q00498; Q00310;
DC	01-FEB-1994 (Rel. 28, Created)
DT	01-OCT-2000 (Rel. 40, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	A-KINASE ANCHOR PROTEIN 12 (A-KINASE ANCHOR PROTEIN 250 KDA) (AKAP12 OR AKAP250).
GN	AKAP12 OR AKAP250.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxId=9606;
GN	[1]

RP SEQUENCE FROM N.A. (ISOFORM 1).
RX TISSUE=Heart;
MEDLINE=97153077; PubMed=9000000;
RA Neuert J.B., Klauack T.M., Langeberg L.K., Scott J.D.;
RT "Gravlin, an autoantigen recognized by serum from myasthenia gravis
patients, is a kinase scaffold protein.";
RL Curr. Biol. 7:52-62(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=98269042; PubMed=9604001;
RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;
RT "Changes of gene expression by lysophosphatidylcholine in vascular
endothelial cells: 12 up-regulated distinct genes including 5 cell
growth-related, 3 thrombosis-related, and 4 others.";
RL J. Biochem. 123:1119-1126(1998).
RN [3]
RP SEQUENCE OF 43-1781 FROM N.A.
RC TISSUE=Umbilical vein endothelial cells;
RA Howditch R.D., Ginsberg M.H.;
RT "Sequence of gravin cDNA isolated from a human umbilical vein
endothelial cell library.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1477-1781 FROM N.A.
RC TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=92395179; PubMed=1522245;
RA Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,
RA Lindstrom J., Ginsberg M.H.;
RT "Molecular cloning and preliminary characterization of a novel
cytoplasmic antigen recognized by myasthenia gravis sera.";
RL J. Clin. Invest. 90:992-999(1992).
CC -1- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR
COMPARTMENTATION OF PROTEIN KINASE (PKA) AND PROTEIN KINASE C
(PKC).
CC CC
CC -1- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL
CYTOSKELETON.
CC CC
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; MAY BE
PRODUCED BY ALTERNATIVE SPLICING.
CC CC
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, CULTURED
FIBROBLASTS AND OSTEOSARCOMA, BUT NOT IN PLATELETS, LEUCOCYTES,
MONOCYTIC CELL LINES OR PERIPHERAL BLOOD CELLS.
CC CC
CC -1- INDICATION: ACTIVATED BY LYSOPHOSPHATIDYLCHOLINE (LYSOPC).
CC CC
CC -1- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE
INVOLVED IN BINDING PKC.
CC CC
CC -1- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY
PATIENTS WITH MYASTHENIA GRAVIS (MG).
CC CC
CC -----
CC CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC CC
CC -----
CC EMBL: U01607; AAC51366.1; -
DR EMBL: AF001504; AAB58938.1; -
DR EMBL: AB003476; BAA1927.1; -
DR EMBL: M96322; AAA35931.1; -
DR MIM: 604698; -
DR PIR: A43922; A43922.
DR InterPro: IPR001573; -
DR InterPro: IPR001899; -
KW Antigen; Alternative splicing.
FT DOMAIN 603 633 AKAP 1.
FT DOMAIN 752 782 AKAP 2.
FT DOMAIN 797 827 AKAP 3.
FT DOMAIN 98 101 POLY-GU.
FT DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).
FT DOMAIN 1540 1553 RII-BINDING (PROBABLE).
FT VARSPLIC MISSING (IN ISOFORM 2).
FT

FT	VARSPLIT	99	106	E => K (IN REF. 2).	E => K (IN REF. 2).
FT	CONFLICT	117	117	NRN -> TPET (IN REF. 2 AND 3).	NRN -> TPET (IN REF. 2 AND 3).
FT	CONFLICT	142	144	Q -> K (IN REF. 2).	Q -> K (IN REF. 2).
FT	CONFLICT	215	215	G -> E (IN REF. 2 AND 3).	G -> E (IN REF. 2 AND 3).
FT	CONFLICT	448	448	R -> G (IN REF. 2 AND 3).	R -> G (IN REF. 2 AND 3).
FT	CONFLICT	694	694	G -> S (IN REF. 2 AND 3).	G -> S (IN REF. 2 AND 3).
FT	CONFLICT	867	867	S -> A (IN REF. 2 AND 3).	S -> A (IN REF. 2 AND 3).
FT	CONFLICT	986	986	E -> EE (IN REF. 3 AND 4).	E -> EE (IN REF. 3 AND 4).
FT	CONFLICT	1530	1530	V -> M (IN REF. 4).	V -> M (IN REF. 4).
FT	CONFLICT	1581	1581	Q -> L (IN REF. 2).	Q -> L (IN REF. 2).
EQ	SEQUENCE	1781	1601	MM; BAB13937379F9EAC0F CRC64;	MM; BAB13937379F9EAC0F CRC64;

Query Match	5.8%	Score 122.5	DB 1	Length 1781
Best Local Similarity	22.0%	Pred. No. 5.9		
Matches 93	Conservative 62	Mismatches 152	Indels 115	Gaps 22

QY	40	PDTV-LYDVLVEDSDEEVLVADPVEVVPARLLPAPAK---PEODSDSDSEGAEEGPAGAPR	95
Db	192	PDTYQVLLTVKKDEBGGAAGAGDHDQDPSLGAEEAASKSEPKOSTERKEE-----	240
QY	96	TLVRRRRRLLDPCGAPVPVYSGKVQSSNLNLFDNSLSLKLCPSEPD--EADLTNCSG	153
Db	241	TLRKQASHAELSP-----PAESGA-----VEECKEEBEEKEPKSKSAE	281
QY	154	SPSEDDALPSSGP-----WRKK--LRKKCEKE--EKKMEFPDQDLSPLQD	166
Db	282	SPTSPVTSSETGSTFKKFFPTQMGAMRKRTSPRKREDEVEASEKKKQDPEK-----VDT	356
QY	197	SSRKSRSKHTALOKLRKVNKRIDDLSCLSPK---QHOSPALDSTDEVLVYEGP----	249
Db	337	EEDGGAELVASERKLVAEQAHPQ--EPAAEHAPRLSAEEKVELPS--EEQVSGSGPSEEK	394
QY	250	VLPOSSRLFTLKICORADLVRLPRMSEPLONVVDHMANHLGVSPNRILLFGESLSPT	309
Db	395	PAPLATEVEFDEKIEVHQGEVAEVAHVSVTERBEO-----KTEVEET	437
QY	310	A--TPSTILKGLVADILDCVVLASSSEATEIYOELRL--YQGEKKOMLEISPSPL	364
Db	438	AGSVPAEELVEMD-----AEPOEAEPARELVKLKETCVSGEDPQOGAD--LSPEKV	487
QY	365	-----RYLMSHYEAMGLSGHKLSFFPDGT---KLSGKFLPADLG---LESGDLI	408
Db	488	LSKPEGVEGVSEVEMSSGSRKAKVQGSPLKFLFTYTGKTKLSGKKQKRGGDSESEHT	547
QY	409	EY 410	
Db	548	QY 549	

RESULT	8	
TCF8_RAT		
ID	TCF8_RAT	STANDARD; PRT; 1109 AA.
AC	Q62947; Q62948;	
DT	01-OCT-2000 (Rel. 40, Created)	
DT	01-OCT-2000 (Rel. 40, Last sequence update)	
DT	01-OCT-2000 (Rel. 40, Last annotation update)	
DE	TRANSCRIPTION FACTOR 8 (ZINC FINGER HOMEODOMAIN ENHANCER-BINDING PROTEIN) (ZfHEF).	
GN	TCF8.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10110;	
RN	(1)	
RP	SEQUENCE OF 6-1109 FROM N.A., AND ALTERNATIVE SPLICING.	
RX	MEDLINE=96365389; PubMed=8769566;	
RA	Cabanillas A.M., Darling D.S.;	
RT	"Alternative splicing gives rise to two isoforms of Zfhef, a zinc finger/homeodomain protein that binds T3-responsive elements."	
RT	DNA Cell Biol. 15:643-651(1996).	
CC	1-1 FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX	

SEQUENCES IN THE IMMUGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES (BY SIMILARITY).

-1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

-1 ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1/ZFHEP-1 (SHOWN HERE) AND 2/ZFHEP-2; ARE PRODUCED BY ALTERNATIVE SPLICING.

-1 SIMILARITY: BELONGS TO DELTA-EFL/ZFH-1 FAMILY OF TWO-HANDED ZINC FINGER/HOMEDOMAIN PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	
DR	EMBL; U51583; AAB17130.1; -;
DR	EMBL; U51584; AAB17131.1; -;
DR	HSSP; P08046; 1AIG.
DR	InterPro: IPR000662; -;
DR	Pfam; PF00096; 2f-C2H2_7.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1_5.
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2_6.
KW	Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW	Homeobox; Repressor; Activator; Metal-binding; Repeat;
KW	Alternative splicing.
FT	ZN_FING 150 173
FT	ZN_FING 180 202
FT	ZN_FING 220 242
FT	ZN_FING 248 272
FT	DNA_BIND 559 618
FT	ZN_FING 881 903
FT	ZN_FING 909 931
FT	ZN_FING 937 958
FT	DOMAIN 968 1109
FT	VARSPLIC 1 198
FO	SEQUENCE 1109 AA; 121626 MW; BEEFE29IC875DDA6 CRC64;

Query Match	Similarity	21.4%	Pred. No. 4.1	DB 1	Length 1109
Best Local	Conservative	57	Mismatches 152	Indels 110	Gaps 19
QY	61	PVEVPARLPAPAKEEDSDSDSEGAAGAPAGAPRTVLRRRRRRLLDGEPVVPVVS-G	119		
Db	535	PAQPPPP---PAPATEKPESSASSAGNGD-----LSPSGPLKNTLSLL	574		
QY	120	KVQSSLLNLPDNNSSILKLCPE--PEDA----DLTNGSSPSEDDALPSCGSPWKKTRK	173		
Db	575	KAYVALNQPSTTEELTKLADSVNLPDLVKKWKEFMQAGQIPGOSLEPPSPDPSGNGTIPA	634		
QY	174	KCEKEKME--EFPDOD-----ISP-LDPPSSRNKSRKHTALOKLREVNKRQ	220		
Db	635	KTEEPQPVVDNGENPEDSTRGOSPLKMTSSPYLPVPGSAINGRSCTSPSPPLNLSANP	694		
QY	221	DLRSLSPKQHOSPALQSTDEEVLVEGVPLEQS--SRLF-----TLKTRCADL	268		
Db	695	QGYSCSVSEGTQEPQVEPLDLSLPKQGGELLERSTVSSVYONSYSVOEPEPLNTSCAR--	752		
QY	269	VRLP-----VMSSEFLOLVVDHMANHLDVSPNRILLFGESELSPTAT---PSTLKLQVA	320		
Db	753	-KEPKDSCVTDSEPVVNVVPPSANPITA-----IPTVAOLPTTVAIDQ	798		
QY	321	DLIDCVVLASSSEATETQELRLRPQGEKHKOML-----EISLSP---DSPKLVLN	368		
Db	799	NSVPC-----LRLANLQKQITILPQVATYTSATVSPAMQEPKVIQIP	841		
QY	369	-SHVEAMGLSGHKLSF---FPDGTKLISGELPADDLGEGDLIE	409		
Db	842	NGNODERQDTSSEGVSDQNDSDCTTPKKTRKEMNMVACDLD	887		

```

RESULT 9
PSC_DROME STANDARD; PRT; 1603 AA.
ID PSC_DROME
AC P35820;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE POSTERIOR SEX COMB PROTEIN.
GN PSC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92018190; PubMed=1833647;
RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT Drosophila genes posterior Sex Combs and Suppressor two of zeste
RT encode proteins with homology to the murine bml-1 oncogene."
RL Nature 353:351-353(1991).
CC -1- FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
CC EXPRESSION PATTERNS OF THE HOMEOYTIC SELECTOR GENES OF THE
CC ANTENNAPEDIA (ANTP-C) AND BITTORAX (BX-C) COMPLEXES, AND HENCE FOR
CC THE MAINTENANCE OF SEGMENTAL DETERMINATION.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59275; CAA41965.1; -
CC DR PIR: S17983; S17983.
CC DR FLYBase: FBgn0005624; Psc.
CC DR InterPro: IPR001841; -.
CC DR Pfam: PF00097; zf-C3HC4; 1.
CC DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
CC KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
CC FT DOMAIN 47 53 POLY-THR.
CC FT DOMAIN 83 88 POLY-THR.
CC FT DOMAIN 91 98 POLY-THR.
CC FT DOMAIN 145 152 POLY-THR.
CC FT DOMAIN 184 202 POLY-SER.
CC FT ZN_FING 265 303 C3HC4-TYPE.
CC FT DOMAIN 642 651 POLY-SER.
CC FT DOMAIN 1066 1069 POLY-GLY.
CC FT DOMAIN 1185 1189 POLY-PRO.
CC FT DOMAIN 1214 1217 POLY-PRO.
CC FT DOMAIN 1391 1396 POLY-PRO.
CC FT DOMAIN 1458 1461 POLY-ALA.
CC FT DOMAIN 1517 1520 POLY-GLY.
CC SQ SEQUENCE 1603 AA; 169999 MW; 77024F409736473 CRC64;

Query Match 5.78; Score 119; DB 1; Length 1603;
Best Local Similarity 18.9%; Pred. No. 8.3;
Matches 68; Conservative 54; Mismatches 123; Indels 114; Gaps 13;

```

```

DB 715 SPERTLNRAITPPSPVQOASAP-----KSKGNLDDSTLMKPPSCMPKRSIASTRKS 769
QY 203 RKHTALQKLRVKNKRQDLRSLSPKROHSPALQSTDEEVL-----VEG 248
DB 770 KEYPKAVSK-----KQKLSPLPTVDFKIRLPVTNGSSGTASPKIEK 812
QY 249 PVLPQSSR---LFTLKIRCRADLVLP-----YMSSE-----LQNVVDHMA- 287
DB 813 PLMPPPAKPMLAPRKLOPSAQFAPPSPIHHAGVOMSAFGNRTPIAKRQPIIPKASR 872
QY 288 -NHGVSPNRIILLF-----GESELSPTATPST 314
DB 873 PNPPANIPDNVNRLLKDAETEIKSIGGSVENNSNAQKPHLYGPKGETYKMPALPAT 931

RESULT 10
MSL1_DROME STANDARD; PRT; 1039 AA.
ID MSL1_DROME
AC P50535;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MALE-SPECIFIC LETHAL-1 PROTEIN.
GN MSL-1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE OF 85-1039 FROM N.A.
RX STRAIN=CANTON-S;
RX MEDLINE=93314941; PubMed=8325488;
RA Palmer M.J., Mergner V.A., Richman R., Manning J.E., Kuroda M.I.,
RA Lucchesi J.C.;
RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster
RT encodes a novel protein that associates with the X chromosome in
RT males."
RT Genetics 134:545-557(1993).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE=95300219; PubMed=7781064;
RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,
RA Kuroda M.I.;
RT "Expression of msl-2 causes assembly of dosage compensation
RT regulators on the X chromosomes and female lethality in Drosophila."
RT Cell 81:867-877(1995).
CC -1- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
CC TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
CC DOSAGE COMPENSATION). MSL-1 IS A PIONEER PROTEIN. MLE, MSL-1 AND
CC MSL-3 ARE CO-LOCALIZED ON THE X CHROMOSOME. EACH OF THE MSL
CC PROTEINS REQUIRES ALL THE OTHER MSLs FOR WILD-TYPE X-CHROMOSOME
CC BINDING.
CC -1- SUBUNIT: MSL-1 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-2.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; MSL-1 IS ASSOCIATED WITH HUNDREDS
CC OF DISCRETE SITES ALONG THE LENGTH OF THE X CHROMOSOME IN MALES
CC AND NOT IN FEMALES, AND IS ALSO ASSOCIATED WITH 10-20 AUTOSOMAL
CC SITES IN FEMALES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: I42514; AAA98918.1; -
CC DR FLYBase: FBgn0005617; msl-1.
CC KW Nuclear protein.
CC FT CONFLICT 188 193 PLPPAA -> HCHLLP (IN REF. 1).
CC FT CONFLICT 492 492 L -> S (IN REF. 1).
CC FT CONFLICT 670 670 I -> M (IN REF. 1).

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 ; Search time 85.95 Seconds
(without alignments)
1063.422 Million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099
Sequence: 1 MAEPLRGKGRPSRGGRARR.....GKELPADLGESGDLIEWG 412

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2803329 seqs, 221847457 residues

Total number of hits satisfying chosen parameters: 2803329

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2099	100.0	412	11	US-08-755-592-6
2	2099	100.0	412	11	US-08-755-592A-6
3	2099	100.0	412	15	US-09-175-254-4
4	1399.5	66.7	419	15	US-09-175-254-3
5	154.5	7.4	1279	1	PCT-US01-04098A-1631
6	154.5	7.4	1301	1	PCT-US01-04098A-3599
7	145.5	6.9	622	20	US-09-641-377-505
8	145.5	6.9	628	20	US-09-641-377-506
9	141	6.7	530	20	US-09-641-377-507
10	133.5	6.4	725	16	US-09-252-991A-23752

11	132	6.3	1708	21	US-09-733-089-18072	Sequence 18072, A
12	132	6.3	1708	22	US-09-816-660-18072	Sequence 18072, A
13	130.5	6.2	1488	21	US-09-733-089-18070	Sequence 18070, A
14	130.5	6.2	1488	22	US-09-816-660-18070	Sequence 18070, A
15	128.5	6.1	755	20	US-09-619-049-240	Sequence 240, App
16	128.5	6.1	755	23	US-60-161-932-1882	Sequence 1882, Ap
17	128.5	6.1	755	23	US-60-167-217-4680	Sequence 4680, Ap
18	128.5	6.1	755	23	US-60-171-622-372	Sequence 372, App
19	128.5	6.1	755	23	US-60-173-464-3782	Sequence 3782, Ap
20	128.5	6.1	755	23	US-60-191-637-4587	Sequence 4587, Ap
21	128.5	6.1	755	23	US-60-191-681-3621	Sequence 3621, Ap
22	128.5	6.1	2024	23	US-60-167-217-5735	Sequence 5735, Ap
23	128.5	6.1	2024	23	US-60-173-464-4625	Sequence 4625, Ap
24	127.5	6.1	1167	23	US-60-167-217-56	Sequence 56, Appl
25	127.5	6.1	1167	23	US-60-173-464-51	Sequence 51, Appl
26	127.5	6.1	1167	23	US-60-191-637-63	Sequence 63, Appl
27	127.5	6.1	1167	23	US-60-191-681-51	Sequence 51, Appl
28	127.5	6.1	1408	23	US-60-167-217-2955	Sequence 2955, Ap
29	127.5	6.1	1408	23	US-60-173-464-2417	Sequence 2417, Ap
30	127.5	6.1	1408	23	US-60-191-637-2910	Sequence 2910, Ap
31	127.5	6.1	1408	23	US-60-191-681-2331	Sequence 2331, Ap
32	127	6.1	442	18	US-09-417-507-40178	Sequence 40178, A
33	127	6.1	1368	20	US-09-619-049-396	Sequence 396, App
34	127	6.1	1368	23	US-60-167-217-7686	Sequence 7686, Ap
35	127	6.1	1368	23	US-60-171-627-633	Sequence 633, App
36	127	6.1	1368	23	US-60-173-464-6151	Sequence 6151, Ap
37	127	6.1	1368	23	US-60-191-637-7599	Sequence 7599, Ap
38	127	6.1	1368	23	US-60-191-681-5928	Sequence 5928, Ap
39	126.5	6.0	676	1	PCT-US01-04703-28	Sequence 28, Appl
40	126.5	6.0	717	1	PCT-US01-04703-29	Sequence 29, Appl
41	125.5	6.0	664	23	US-60-212-664-602	Sequence 602, App
42	124.5	5.9	518	1	PCT-US00-05883-862	Sequence 862, App
43	124.5	5.9	555	1	PCT-US00-26524B-5837	Sequence 5837, App
44	124.5	5.9	1779	23	US-60-245-228-307	Sequence 307, App
45	124	5.9	1191	1	PCT-US01-04098A-3865	Sequence 3865, Ap

ALIGNMENTS

RESULT 1

US-08-755-592-6

Sequence 6, Application US/08755592

GENERAL INFORMATION:

APPLICANT: Glimcher, Laurie H.

APPLICANT: Ho, I-Cheng.

TITLE OF INVENTION: Methods for Regulating T Cell Subsets by

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESSES:

ADDRESSER: LAHIVE & COCKFIELD

STREET: 60 State Street, suite 510

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/755,592

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/636,602

FILING DATE: 23-APR-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: DeConti, Giulio A., Jr.

REGISTRATION NUMBER: 31,503

REFERENCE/DOCKET NUMBER: HUT-021CP


```

OY      335 TETSOELRLROGGEKHOMLEISLSPSPKVLMSHYEAGMLSGHKLSPFDGOTKLSCG 394
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      481 PKKQ-----KKRAKEKPELLE--KKKDSPOEAKKEEKKKKAAPE-----E 521
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      395 ELPADLGLLE 403
          | | | | : :
Db      522 ETAPKLGVK 530

```

```

RESULT      8
US-09-641-377-506
: Sequence 506, Application US/09641377
: GENERAL INFORMATION:
: APPLICANT: WIEMANN, STEFAN
: APPLICANT: GASSENHUBER, JOHANN
: APPLICANT: TAMPE, JENS
: TITLE OF INVENTION: HUMAN DNA SEQUENCES
: FILE REFERENCE: 087100/0106
: CURRENT APPLICATION NUMBER: US/09/641.377
: CURRENT FILING DATE: 2000-08-18
: PRIOR APPLICATION NUMBER: 60/149,499
: PRIOR FILING DATE: 1999-08-18
: PRIOR APPLICATION NUMBER: 60/156,503
: PRIOR FILING DATE: 1999-09-28
: NUMBER OF SEQ ID NOS: 1793
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 506
: LENGTH: 628
: TYPE: PRT
: ORGANISM: Unknown Organism
: FEATURE:
: OTHER INFORMATION: Description of Unknown Organism: Database comparison
: OTHER INFORMATION: sequence used for homology purposes
US-09-641-377-506

```

Query Match	6.9%	Score 145.5	DB 20	Length 628
Best Local Similarity	22.1%	Pred. No. 0.042		
Matches 95	Conservative 63	Mismatches 166	Indels 105	Gaps 20

```

OY 335 TETSOEELRLRQGGKEKHQMEIISLSPSPGLKVLMSHVEAMGSLGNHLSLSPFPOTKLSG 394
Db 477 PKEQ-----KKAEKEKPLT--KPKDSPGAKEKEKEKKAAP-----E 517
OY 395 ELPADLGLT 403
Db 518 ETAPKLGVR 526

```

```

RESULT      9
US-09-641-377-507
; Sequence 507, Application US/09641377
; GENERAL INFORMATION:
; APPLICANT: WIEMANN, STEFAN
; APPLICANT: GASSENHÜBER, JOHANN
; APPLICANT: TAMPE, JENS
; TITLE OF INVENTION: HUMAN DNA SEQUENCES
; FILE REFERENCE: 087100/0106
; CURRENT APPLICATION NUMBER: US/09/641,377
; CURRENT FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/149,499
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/156,503
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 1793
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 507
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Database comparison
; US-09-641-377-507

```

Query Match	6.7%;	Score 141;	DB 20;	Length 530;
Best Local Similarly	22.9%;	Pred. No. 0.073;		
Matches 89;	Conservative 57;	Mismatches 151;	Indels 92;	Gaps 19;

? CURRENT FILING DATE: 1999-02-18
 ? PRIOR APPLICATION NUMBER: US 60/074,788
 ? PRIOR FILING DATE: 1998-02-18
 ? PRIOR APPLICATION NUMBER: US 60/094,190

US-09-619-049-240

Query Match 6.1%; Score 128.5; DB 20; Length 755;

Best Local Similarity 24.1%; Pred. No. 1.1; Matches 77; Conservative 48; Mismatches 122; Indels 73; Gaps 16;

```
QY 71 APAKPEQSDSDSGAEGPAGAPRT--LVRRRRRRL-----DPCGAPVVPVYSGKVQ 122
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 41 AEAKPR-----DSAPAKSQENSSVTPOQTKRKPRKRLSLAEDQNEPEAPDATANNTSAR 95
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 123 SSLNLIPDNGSLKLCPSPEDEADLTNSGSSP-SEDDALPGSGPWR-KKLRKCKEKEEK 180
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 96 QSTFRV--SNSQLAIADENHNSTASLMPPPVPVPSADTTLGSGRPQRAKL-----KTEK 148
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 181 KMEFPDQDISPLPQPSRNKSRK-HTDALOKLR-EVKNRLQDLRSC-----LSP 228
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 149 LKLE-----PSLNKKMRKRPSSSELVKVESEQRYSQPNSTTSAQALEENKLAE 197
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 229 KOHSPALQSTDEVLVEGVPVLPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMAN 288
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 198 PELEPPAETAEOQKPEPEASVTEDEVNTTKTLRYKVARE--KLSTEAVPLTNVAVS--TAN 254
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 289 HLGVSFNRILLGEGSELSPTATPSTLKLGVADIIDCVVLASSSEATETSOELRLRVQOK 348
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 255 VTTVS-----SVTTEAARP-----DDTVASNTTSTEVSKVKVKKKKKDV 293
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 349 EKHOMEI-----SLSPDSP 364
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 294 ESHRPKIVERPSDLDKSSPV 313
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

Search completed: September 11, 2001, 08:55:30
Job time: 424 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 ; Search time 225.87 Seconds
(without alignments)
46.594 Million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099
Sequence: 1 MAEPLRGGRGPRSGRGARR.....GKELPADLGESGDLIEWWG 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 204157 seqs, 25543879 residues

Total number of hits satisfying chosen parameters: 204157

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA.New.*
1: /cgn2_6/prodata1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	952	45.4	235	US-09-758-462-1272	Sequence 1272, Ap
2	926	44.1	248	PCT-US01-14827-14233	Sequence 14233, A
3	635	30.3	259	US-09-758-472-5521	Sequence 5521, Ap
4	250	11.9	91	PCT-US01-14827-14232	Sequence 14232, A
5	201.5	9.6	146	US-09-758-472-5605	Sequence 5605, Ap
6	164	7.8	78	US-09-758-472-9144	Sequence 9144, Ap
7	146	7.0	716	PCT-US01-14827-8086	Sequence 8086, Ap
8	122.5	5.8	1003	PCT-US01-14827-14803	Sequence 14803, A
9	122.5	5.8	1781	US-09-738-877-3	Sequence 3, App1
10	120.5	5.7	565	US-09-765-272-218	Sequence 218, App
11	118	5.6	227	US-09-758-440-512	Sequence 512, App
12	118	5.6	576	US-09-882-636-2	Sequence 2, App1
13	117	5.6	1451	PCT-US01-14827-9935	Sequence 9935, Ap
14	116.5	5.6	122	PCT-US01-14827-8085	Sequence 8085, Ap
15	115.5	5.5	458	US-09-758-438-749	Sequence 749, App
16	115.5	5.5	1342	PCT-US01-14827-13151	Sequence 13151, A
17	114.5	5.5	92	US-09-864-761-36307	Sequence 36307, A
18	113	5.4	554	PCT-US01-14827-11206	Sequence 11206, A
19	113	5.4	3439	PCT-US01-14827-11861	Sequence 11861, A
20	112.5	5.4	95	US-09-834-366-20964	Sequence 20964, A
21	112.5	5.4	95	US-09-834-366-26272	Sequence 26272, A
22	112.5	5.4	95	US-09-538-092-1221	Sequence 1221, App
23	112.5	5.4	144	PCT-US01-18569-2835	Sequence 2835, App
24	111	5.3	605	US-09-243-560A-2	Sequence 2, App1
25	110.5	5.3	428	US-09-893-348-18	Sequence 18, App1
26	110.5	5.3	462	US-09-506-514-4	Sequence 4, App1
27	109.5	5.2	362	PCT-US01-14827-11292	Sequence 11292, A

28	109.5	5.2	710	US-09-079-812E-2	Sequence 2, App1
29	109	5.2	567	PCT-US01-16450-2705	Sequence 2705, Ap
30	109	5.2	1364	PCT-US01-14827-11860	Sequence 11860, A
31	108.5	5.2	1085	PCT-US01-14827-10316	Sequence 10316, A
32	108.5	5.2	1991	PCT-US01-14827-13152	Sequence 13152, A
33	108	5.1	129	PCT-US01-08656-7198	Sequence 7198, Ap
34	108	5.1	143	US-09-760-469-1289	Sequence 1289, Ap
35	108	5.1	549	US-09-764-868-629	Sequence 629, App
36	108	5.1	675	US-09-738-626-5917	Sequence 5917, App
37	107.5	5.1	138	US-09-757-029-186	Sequence 186, App
38	107	5.1	2665	US-09-864-761-34248	Sequence 34248, A
39	106.5	5.1	380	PCT-US01-14827-10911	Sequence 10911, A
40	106.5	5.1	437	US-09-864-761-36083	Sequence 36083, A
41	106.5	5.1	723	US-09-368-831-5	Sequence 5, App1
42	106	5.1	350	US-09-758-437-608	Sequence 608, App
43	105.5	5.0	485	PCT-US01-14827-13477	Sequence 13477, A
44	105.5	5.0	734	PCT-US01-14827-13022	Sequence 13022, A
45	105.5	5.0	829	PCT-US01-14827-12121	Sequence 12121, A

ALIGNMENTS

```
RESULT 1
US-09-758-462-1272
; Sequence 1272, Application US/09758462
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM007
; CURRENT APPLICATION NUMBER: US/09/758,462
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1734
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1272
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-758-462-1272

Query Match          45.4% Score 952; DB 5; Length 235;
Best Local Similarity 81.2% Pred. No. 3.8e-52;
Matches 190; Conservative 13; Mismatches 31; Indels 0; Gaps 0;

QY 179 EKKKEFPDODISPLPOPSSRNKSRKHTALQKREYNKRLODLRSCLSPKOHSPALQS 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 EKKTEFLDLNDSLPSPSPRTKSRTHTRALKKSEVNNKRLQDLRSCLSPPPGQSGQG 61

QY 239 TDDEVVLVEGVLPOSSRLFTLTKRCRADLVRLPVRMSEPLQNVDMANHLGVSPNRIL 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 QEDDEVVLVEGVLTPETRLPLKTRCRADLVRLPLRMSEPLQSVDMATHLGVSPNRIL 121

QY 299 LIREESLSPAMPSTKLGVADIIDCVLASSSEARETSQELRLVROGKEMHLEISL 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 LIRETELSPAPPRTLKLGADIIDCVLSSSEARETSQQLRVQGEKHQTLVLSL 181

QY 359 SPDPSLVKLVMSHVEAMGLSGHKLSFFFDSTKLSGKELPADLGESGDLIEWWG 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 SRDSPKLTMSHVEAMGLSGRKLSFFFDSTKLSGRLPADLGESGDLIEWWG 235

RESULT 2
PCT-US01-14827-14233
; Sequence 14233, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
```



```

? LOCATION (31)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION (126)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION (129)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION (135)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION (136)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-472-5605

```

Query Match	9.68	Score 201.5	DB 5	Length 146
Best Local Similarity	41.78	Pred. No. 5.6e-06		
Matches 58	Conservative 9	Mismatches 57	Indels 15	Gaps 3

```

oy      4 PLKGGPKRSKSGRGARAK-----GAGKCPCPRARDSPARLLP---DYLVDLSDSDEE   54
          | | | | | | | | | | : |||||
Db      1 PTRCGGPGCEGXGPLVRRXRCRPRGSGRLGRXGPASSGPAVSIPGHAGKMSVDLVTDSEF 60

```

```

QY      55 VLEY-----ADPVEYVVARLRPAKPEQSDSDSEGAAGPAGAPRTLVARRRRRLIDP 108
      :|||      |||      |      :|||      |||      :|||
Db      61 ILEVATARGAADEVETPEPPGVPVASRDNSNDSSEGEDRRPADPRTGSADGRLVLDP 120

```

```
QY      109  GEAPVPVYSGKVQSSLNL 127
          ||||: || |||: |
Db      121  GEAPLPVYXSGKLKXXFAL 139
```

6
 US-09-758-472-9144
 : Sequence 9144, Application US/09758472
 : GENERAL INFORMATION:
 : APPLICANT: kosen et. al.
 : TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 : FILE REFERENCE: PH001
 : CURRENT APPLICATION NUMBER: US/09/758, 472
 : CURRENT FILING DATE: 2001-01-11
 : PRIOR APPLICATION NUMBER: 60/179, 065
 : PRIOR FILING DATE: 2000-01-31
 : PRIOR APPLICATION NUMBER: 60/180, 628
 : PRIOR FILING DATE: 2000-02-04
 : NUMBER OF SEQ. ID NOS: 9632
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ. ID NO 9144
 :
 : LENGTH: 78
 :
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : US-09-758-472-9144

Query Match	7.8%	Score 164	DB 5	Length 78
Best Local Similarity	50.6%	Pred. No. 0.00048		
Matches 39	Conservative	9	Mismatches 27	Indels 2
			Gaps	2

QY 109 GEAPVVPVYSGKVQSSLNLPDSSLLKLCPSPEDEADLTNSGSSPSEDDALPGSPWR 168
 || : | : ||| |||| : ||| : | : ||| :
 Db 3 GEPRLPFPTKGRFKSSRLIPDDLSLKLTPPGDEEEAEALADSSGLIYHEGSPSP -GSPWK 61

```
QY 169 KKLRRKCEKEEKKMEEF 185
    ||| | : |||| | |
Db 62 TKLR TK-DKEEKKKTEF 77
```

```

RESULT      7
PCT-US01-14827-8086
; Sequence 8086, Application PC/TUS0114827
; GENERAL INFORMATION:

```

```

? APPLICANT: Hyseng, Inc
? TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
? FILE REFERENCE: 21272-104
? CURRENT APPLICATION NUMBER: PCT/US01/14827
? CURRENT FILING DATE: 2001-05-16
? PRIOR APPLICATION NUMBER: 09/577,408
? PRIOR FILING DATE: 2000-05-18
? NUMBER OF SEQ ID NOS: 16102
? SOFTWARE: Custom
? SEQ ID NO 8086
? LENGTH: 716
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: DOMAIN
? LOCATION: (654)..(706)
? OTHER INFORMATION: Ubiquitin domain proteins domain identified by EMBLPIR.
? OTHER INFORMATION: accession number BI00299, p-value=4.420e-18, raw score of 28.8
? NAME/KEY: DOMAIN
? LOCATION: (401)..(608)
? OTHER INFORMATION: Reverse transcriptase domain identified by Pfam, accession
? OTHER INFORMATION: name rvt, E-value=2.2e-11, Pfam score of 51.2
PCT-US01-14827-8086

```

Query Match	7.0%:	Score 146:	DB 1:	Length 716:
Best Local Similarity	18.7%:	Pred. NO. 0.15:		
Matches 89;	Conservative 73;	Mismatches 173;	Indels 142;	Gaps 17:

```

Qy 48 VSDSEEVLEVADPEVFPVARLPAPAKPEQ--DSDSDSCGAEGPAGAPRLVRRRRR 104
    ::|  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 261 INESMSWFFFEINKIDRPLARLKKKKKRKNPIDAIKNDGDTTDPTEIQTTIREYVKH 320

```

```

QY      105 LDDPEAPVYVYSGKVQ--SLNLIPINSSLLKLCPSPEED-EADLTNS-----GSSP 155
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      321 L-----YANKLENLSEDKFLDTYTLPRNQEEVESLNRPITGSEIVAIINSLP 365

```

```

Oy 156 SEDDALPSG-----SPWRKILRKCKEEKKNEFPDDDISLPQPSSRKKS 202
      ::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 370 TKKSPGPDGTTTESYQRYKKELVPLLKLSQSIKEGILPNSFYEASIIILPKP-GRDTT 428

```

```

Oy 203 RKHT-----BALQKL-----REVN---- 216
      :|      : :||      :|
Db 429 KKENERPISLNIADAKILANKILANQIQIHRIKLTHNDQVGFIRGMQGWFNICKSINVQH 488

```

```

Qy 217 -----KRLQDLRSC-----LSPKQHQSPA-----LQS 238
      |      |      |      |      |
      ::: |      |      |      |
Db 489 RNRANKNHMIISIDAEKAFDEIQQCFMKTINKLEPSNHTALAGVGHQEDSIALAYD 548

```

OY 239 TDDEVVVLEGVVLPOSSRLFTIKRCRADLVLRVPVMSEPLQNVDMANHLGVSPPRI- 237
| : | | | : | : | : | : | :
DB 549 LDENHNESQLAAPDSHIIQ---SCRKNQLSAPFSAGSAVHAAD-MHSCMASSPYRVR 604

```

Qy 298 ---LTLFGESLSPYATPSTLKGADIIDCVVLASSSEATETSQGLRLRVQGREKHOML 354
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 605 TRHLCEAAAEETPVFTMANEK-----PTEEKXTENNHNHLKVAQD-GSVV 651

```

```
QY      355 EISLSPDPLKVLMSHEAMGLSGHKLSFFPDGTKLSKELPADGLSEGDLEVM 4111
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      652 QFKIKRQPLSKLMKAYCEPCRLSMKQIRFRFGQPISTGDKPAQLEMEDEDTIDVF 708
```

```

RESULT 8
PCT-US01-14827-14803
; Sequence 14803, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
;

```

```

; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 14803
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(1003)
; OTHER INFORMATION: xaa = X or * as defined in Table 2
PCT-US01-14827-14803
```

```

Query Match          5.8%; Score 122.5; DB 1; Length 1003;
Best Local Similarity 26.6%; Pred. No. 6.7;
Matches 54; Conservative 19; Mismatches 77; Indels 53; Gaps 9;
```

```

QY 4 PLRGGRPSRSGRGARR---ARGARGCPRARQSPARLIPDVLVSDSDSEVLEVA 59
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 471 PARGXGPG-XGGRREPEPARARGQA--AGTEPAR-VQALQRLDLDPED----- 520
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 60 DPEVVPARLPA-PAKEQSDSDSEGAEGPAGAPRTLYRRRRLLDGEAPVVPYS 118
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 521 -----ARLTSTPKSRRTAASDPACSLPGGSR----- 550
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 119 GKVOSSILIPDNSILKLC-----PSEPEDADLTNSGSSPEDALPSPGPPRRKLRK 173
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 551 GVPQAMAVOLPESERPPLAACHHSTRPPDPKHAACSLLSGSG---PVLPRSPVPPRASP 607
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 174 KCEKEKKMEEPDODISPLPP 196
    |||  |||  |||  |||  |||  |||  |||  |||
DB 608 KAQLXVQSQMSPPPGIPOLPP 630
    |||  |||  |||  |||  |||  |||  |||  |||
```

```

RESULT 9
US-09-738-877-3
; Sequence 3, Application US/09738877
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan
; APPLICANT: Weiss, Stephen J.
; APPLICANT: Glynn, Richard
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET
; FILE REFERENCE: A-69806/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/738,877
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,425
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT/US 00/22061
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-877-3
```

```

Query Match          5.8%; Score 122.5; DB 5; Length 1781;
Best Local Similarity 22.0%; Pred. No. 15;
Matches 93; Conservative 62; Mismatches 152; Indels 115; Gaps 22;
```

```

QY 154 SPSEDDALPSPGSP-----WRKK--LRKKCKE-----EKKMEFPDODISPLPP 196
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 282 SPSPVPTSETGTFPKKFTTQAGMAGWRKKTSPFRKPKDEVEVASEKKKEQEPK-----VDT 336
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 197 SSRNRSKRHTALOKLREVNKRLODLRSCSPK---OHQSPALQSTDDDEVVLEGP--- 249
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 337 EEOGKAIVASEKLTASBQAHQ--EPASAHPEPRLSAYEKVELPS--EEOVSGSQGSEEK 394
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 250 VLQSSRLFLTKIRCADLVRLVRYRMSEPLQNVVDHANLGVSPRILLPSESELSPT 309
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 395 PAPLATEVFDEKIEVHOEEVAEVHSTVERTEQ-----KTEVEET 437
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 310 A--TPSTLKGVDIIDCVVLASSSEATFTSQELRLR---VQGEKHQMLEISLSPSPL 364
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 438 AGSVPAELVGM-----AEQGEAPRAKELVKLKECTVSEDDPTQGD--LSFDEKV 487
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 365 -----KVLMSHYEAMGLSGHKLSPFFDGT---KISGRELADLG--LESGLTI 408
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 488 LSKPPEGVSEVEMLSSQERMKVQGSPLKLTFTSTGLKTLGKQKQKRGCGDESEGEHT 547
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 409 EV 410
    |||  |||  |||  |||  |||  |||  |||  |||
DB 548 QV 549
    |||  |||  |||  |||  |||  |||  |||  |||

RESULT 10
US-09-765-272-218
; Sequence 218, Application US/09765272
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 218:
US-09-765-272-218
```

```

Query Match          5.7%; Score 120.5; DB 5; Length 565;
Best Local Similarity 20.9%; Pred. No. 4;
Matches 81; Conservative 66; Mismatches 156; Indels 85; Gaps 19;
```

QY 29 PRAROSP-----ARLPDTVLVDLVSDSEVLEVADPVEV--PVARLP--APAKPE-Q 77
Db 34 PKTESPEKEPEKSEVKPTDPTLPEVEGKEDSAPAPVEVEGVESEKPEKVAVKPESQ 93
QY 78 DSDSDSGAAGPAGARPLVRRRRRLLDGEPAPVYVYSGKVQSS--LNLIPDNSSL 134
Db 94 PSDRPAEESKVEQAGEP--VAPREDEKAPVEPEKQPEAPEEKEAVEETPKOESTPTDKAE 152
QY 135 LKICP-----SEPEDEADLTNCSG--SPSDDALPSSGSPWRKK 170
Db 153 ETVPKREYVNOSTEQKVEPAPVEKOTEPTEEPKVEQAGEPAPVAPREDEQAPT--APVEPE 211
QY 171 LRKCKEKKKEEPPDDISPLPEOPSSRNKSRKHTALOKLREVNKRLDRLNSCLSPKQ 230
Db 212 KQPEVPEEKAVEB-----TPKPEDKIGIGTKPEVPDK--SELNMOI--DKASSVSPDT 261
QY 231 HQSALOSTDEVLVGPVLPQSSRLFTLKIRCADLVRLPVMSF--PLQNVVD---- 284
Db 262 YSTASYNAL-----GPVLETAKGYA-----SEPVKQPFVNSETNKLTAKIDALNV 307
QY 285 ---HMANHLGVSPRILLFGESELSPTAPPSITKLGVADIDCVVLASSEATETSQEL 341
Db 308 DKTELNTIADAKTKVKEHSDBRMQNLQTEVYTAKEVYAAANTD---AKQSEVNEAVEK 363
QY 342 RLRYQKEKHOMLEISLSPDSPILKVLMS 369
Db 364 TATIE-----KLVELS---EKPILTLS 383

RESULT 11
US-09-758-440-512
; Sequence 512, Application us/09758440
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM015
; CURRENT APPLICATION NUMBER: US/09/758,440
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 832
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 512
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-758-440-512

Query Match 5.6%; Score 118; DB 5; Length 227;
Best Local Similarity 25.3%; Pred. No. 1.5;
Matches 62; Conservative 40; Mismatches 91; Indels 52; Gaps 11;

QY 1 MAEPLRGGRSGRGARARARAGRCPRARQSPARLIPDTVLVDVSDS--DEVLEV 58
Db 8 VAVPLAGGQEGSPGG--GRKSRGT---TWYKRRKGRVYIDSDTDESGSDENLDQELLST 62
QY 59 A-----DPEVAVARLPAPAKPEODSDSDEGACGAPGAPRTVLVRRRRRLLDGGA 111
Db 63 AKRRRSSEKEPEPVSG--PAASSDSETSDDDEWT-----FGSNKNKKKRAKKEKG-- 115
QY 112 PVVYVYSGVQSSINLIPDNSSLKICPSEPEDEADLTNCSGSPSEDDALPSSGSPWRKK 171
Db 116 -TMKQANKTAASSGSSDKD--SSAESSAPEREGEVSDSDNSSSSSSSDSDS----- 162
QY 172 RKCKEKKKEEPPD-----QDISPLPQSSRNKSRKHTALOKLREVNKRLD 221
Db 163 -----SSEDEPHDYGEDLMGDEEDRAKLEQMTKERQELFNRIEK--REVLKRFE 214
QY 222 LRSC 226

Db 215 IKKKL 219
RESULT 12
US-09-882-636-2
; Sequence 2, Application US/09882636
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth
; APPLICANT: Lawrence, David, A
; APPLICANT: Roy, Margaret, Ann
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2509R1C1
; CURRENT APPLICATION NUMBER: US/09/882,636
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/107,783
; PRIOR FILING DATE: November 10, 1998
; PRIOR APPLICATION NUMBER: 60/088,742
; PRIOR FILING DATE: June 10, 1998
; PRIOR APPLICATION NUMBER: 60/086,414
; PRIOR FILING DATE: May 22, 1998
; PRIOR APPLICATION NUMBER: 60/083,500
; PRIOR FILING DATE: April 29, 1998
; PRIOR APPLICATION NUMBER: 60/082,767
; PRIOR FILING DATE: April 23, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/066,772
; PRIOR FILING DATE: November 24, 1997
; PRIOR APPLICATION NUMBER: 60/032,705
; PRIOR FILING DATE: December 12, 1996
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: 09/709,238
; PRIOR FILING DATE: November 8, 2000
; PRIOR APPLICATION NUMBER: 09/664,610
; PRIOR FILING DATE: September 18, 2000
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: September 18, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: August 23, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: May 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/13705
; PRIOR FILING DATE: May 17, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05004
; PRIOR FILING DATE: February 24, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04341
; PRIOR FILING DATE: February 18, 2000

```

; PRIOR APPLICATION NUMBER: PCT/US00/04342
; PRIOR FILING DATE: February 18, 2000
; PRIOR APPLICATION NUMBER: 09/480,284
; PRIOR FILING DATE: January 10, 2000
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: December 2, 1999
; PRIOR APPLICATION NUMBER: 09/423,844
; PRIOR FILING DATE: November 12, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: November 30, 1999
; PRIOR APPLICATION NUMBER: 09/403,297
; PRIOR FILING DATE: October 18, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/20111
; PRIOR FILING DATE: September 1, 1999
; PRIOR APPLICATION NUMBER: 09/380,137
; PRIOR FILING DATE: August 25, 1999
; PRIOR APPLICATION NUMBER: 09/380,138
; PRIOR FILING DATE: August 25, 1999
; PRIOR APPLICATION NUMBER: 09/380,139
; PRIOR FILING DATE: August 25, 1999
; PRIOR APPLICATION NUMBER: 09/367,206
; PRIOR FILING DATE: August 9, 1999
; PRIOR APPLICATION NUMBER: 09/369,028
; PRIOR FILING DATE: August 4, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 2, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/08847
; PRIOR FILING DATE: April 23, 1999
; PRIOR APPLICATION NUMBER: 09/298,404
; PRIOR FILING DATE: April 23, 1999
; PRIOR APPLICATION NUMBER: 09/284,291
; PRIOR FILING DATE: April 12, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/05028
; PRIOR FILING DATE: March 8, 1999
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/00106
; PRIOR FILING DATE: January 5, 1999
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: 09/065,275
; PRIOR FILING DATE: April 23, 1998
; PRIOR APPLICATION NUMBER: 08/987,902
; PRIOR FILING DATE: December 10, 1997
; PRIOR APPLICATION NUMBER: PCT/US97/22278
; PRIOR FILING DATE: December 5, 1997
; NUMBER OF SEQ ID NOS: 113
; SEQ ID NO 2
; LENGTH: 576
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-882-636-2

Query Match          5.6%; Score 118; DB 5; Length 576;
Best Local Similarity 21.8%; Pred. No. 5.8;
Matches 90; Conservative 49; Mismatches 148; Indels 126; Gaps 18;

QY 3 EPLRGGRPSRGCGARARGRCRCRARGSPARLLPDTVLVYVSDSDSEVLEVADPV 62
    |||||
Db 137 EPLRKAR-----KMSNSQPADLAHMGSRSDPAGMEASTWIPALPRTS-----SDPV 183
```

```

QY 63 EVPYARLPAPAKPEODS--DSDSGAEGPAGAPRTLVRRRRRLDPGEAPVVPYSGK 120
    : |||
Db 184 ---LTKAPALGYVADSLRASDGLAKAPTKPRT-----PSEFLPDAISRP--PTY--- 231

QY 121 VQSSLANLIPNSSLKLCFSEPEDEADLTNGSSPSEDDALPGSGPWKRRLKCKEKEK 180
    : |||
Db 232 -----CELVPVPSVQ-----GTSFQSCPEPE-APWME--AEDEEEN 268

QY 181 KMEFPDODISPLPQPSRRKSKHTALOKLREVNKRLODDLNSCLSPKHOQ--SPALOS 238
    : |||
Db 269 RCFTRPQAEISFCP-----HDAPSCLGPQNRPLEPQVLH 303

QY 239 TDDEVVLVEGPVL---PQSSRLFTKIRCR-----DVLVRPVRM 275
    : |||
Db 304 T-----LRGLFLEHNGSTALHLLVDCATGLGTVTRQRCNGMVCSSGELLETLPHGH 357

QY 276 SEPLQNVYDMANHL-----GVSPNRILLFGESELSPYAPPTSL--KLGVADII 323
    : |||
Db 358 HLRLELERRQTLALAGALVAGCGPLEERRAALRGLVELALALRGAAGDPLGLAAM 417

QY 324 DCVYLASSFATERSOELRLVQCKEKHOMLETSLSPSPLKYLMSHYEAMG 376
    : |||
Db 418 GALIMPQVSRLEHTWRQLR-----RSHTEALAFEOELKPLMRALDEGAG 462

RESULT 13
PCT-US01-14827-9935
; Sequence 9935, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 9935
; LENGTH: 1451
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (925)..(980)
; OTHER INFORMATION: RNA POLYMERASE NUCLEOCAPSID ALPHA SUBUNIT RNA-D domain
; OTHER INFORMATION: Identified by eMATRIK, accession number PD02316B, p-value=8.51
PCT-US01-14827-9935

Query Match          5.6%; Score 117; DB 1; Length 1451;
Best Local Similarity 20.7%; Pred. No. 25;
Matches 103; Conservative 72; Mismatches 167; Indels 156; Gaps 25;

QY 42 TVLVYVSDSDSEVLEVADPEVP-----VARLPAPAKPEODSDSDSGAEGPAGAPRT 96
    : |||
Db 812 TLEPLFTSESKIF--SSHQIQPOLQDRLLRISQLIOPQD--NLKALQEOIATQREA 866

QY 97 LV-----RRRRRLDPGEAPVVPYSGKVQ-----SSLMLIPDN--SLKLCFSEPEDE 145
    : |||
Db 867 IILARQAREELHJOSE-----WEGRISPEQYDVSPLVQHSFASLPLESESENOE 920

QY 146 A-----DLTNGSS-----PSED-----DALPGSPWRK 169
    : |||
Db 921 PCSINSDNIVSGSHSETPLPDGLGLSHLVLPQDDNLILAEHLAQTFLPSIEKTK 980

QY 170 K--LRKCKEKEKKMEEPDQ-----DISPLPQPSRRKKS--RKHTEA-----LQKLEVN 216
    : |||
Db 981 ELVLSKCKEKEKYSEHFIQSHHGDLQALQOQDLDTQKKAIRISQVQGEELLQRLSELE 1040

QY 217 KRLQDLRSCIS-----PKQHOQ--PALQSTQDEVVLVEGP 249
```



```

Db 1041 KRVSSEYCVSSSEYVQVPAVDSERTQKSFPTKSDTLPSSHREIPLQ---DRLISLSKP 1097
QY 250 VLPOSSRLFTLKIRGRADV-----RLPYRMSEPLQNVYDHANHLGVSPN 295
Db 1098 ILPOODNM-TAQDLAQREVMYSYKEKPOEELSLNQRKLNKSESHHTPLSLFKETPHS 1156
QY 296 RILLFESELSPT-----APPSTLKL-GVADIIDCVVLASSSEATEETSQE-LR 342
Db 1157 FIPLPFAEAKPKSTCYELSSQNEHAAPPSPNVIPGFQD-----RLSPSQSVLTPQODNLG 1211
QY 343 LRVGKREKHOMLEISLSPDSELYK-----LSMHEEMGSLSGHLSFFPGTKISG 393
Db 1212 LQKDLQREVLHYSQKAEKTLVQRTALQOQIOTKHHE-----TLKDFRDSQISK 1263
QY 394 KELPADLGESGDLIEVW 411
Db 1264 PTVENDLKTQKMGOLRDM 1281

```

```

RESULT 14
PCT-US01-14827-8085
; Sequence 8085, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 8085
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (62)..(114)
; OTHER INFORMATION: ubiquitin domain proteins domain identified by eMATRIX.
; OTHER INFORMATION: accession number BL00299, P-value=2.250e-25, raw score of 28.84
PCT-US01-14827-8085

```

```

Query Match 5.6%; Score 116.5; DB 1; Length 122;
Best Local Similarity 28.4%; Pred. No. 0.79; Mismatches 42; Indels 15; Gaps 2;
Matches 31; Conservative 21;

```

```

QY 303 ESELPTATPTSTLKGADIIDCVVLASSSEATEETSQELRLRVOGKEKHOMLEISLSPDS 362
Db 23 EAARSPTGSGFTSREGV-----KTENNNDHINKLVAGOD-GSVVQFKIKRHT 67
QY 363 PLKVLMSHYEAMGLSGHLSFFDGTKLSGKELPADLGESGDLIEVW 411
Db 68 PLSKIMRAYCEROGLSMKQIRFRFDGQPINETDPPAOLEMEDEDTIDVF 116

```

```

RESULT 15
US-09-758-438-749
; Sequence 749, Application US/09758438
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM016
; CURRENT APPLICATION NUMBER: US/09/758,438
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1330
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 749
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (110)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (418)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-438-749

```

```

Query Match 5.5%; Score 115.5; DB 5; Length 458;
Best Local Similarity 25.6%; Pred. No. 6;
Matches 63; Conservative 24; Mismatches 84; Indels 75; Gaps 11;

```

```

QY 34 SPARLIPTVLVLDVSDDEVLEVADPEVPVAPRLPAPAKPEDSDSDSGAAGPAGA 93
Db 237 SSSELSPDAVEKAGKSSNQ---SISPVLDVAVPTPERER-----SSASSPKMDGL 286
QY 94 PRTIVRRRRRLDPGEAPVVPVYSGKVQSSLNLPDMSLT-----LKLCPSEPD 144
Db 287 PRTISRER-----SGSSPGLRDSG-----TPSRHSLSGSSPGMKDIPRTSRGRS 333
QY 145 EADLTNCGSSPSEDDALP-----SGSPWRKKLRKCC--EKEKKMEFPDQ----- 188
Db 334 ECD-----SSP-EPKALPQTPRPRSRSPSPSLNNKCLTPQREERSGSSSVQKTVARTP 387
QY 189 -----DISPLPOPSSR-----NKSRRKHTALQKLRVYNNKRLQDLRGLS 227
Db 388 LGQSRSGSSQELDVKPSASPQENSESDSSXDSKAKTPTPLRQSRSGSSPEVDSKRLS 447
QY 228 PKHQQS 233
Db 448 PRSRRS 453

```

```

Search completed: September 11, 2001, 08:53:51
Job time: 325 sec

```



```

Db 121 VOSSLNLIIPDNSSLLKLCPESEDEADLTNSGSSPSSEDDALPGSPWRKKLRKCKEKEEK 180
Qy 181 KMEFPPODISLPPOSSRNKSRKTEALOKLREVNKRLDOLRSCLSPKOHOSPALOSTD 240
    |||||
Db 181 KMEFPPODISLPPOSSRNKSRKTEALOKLREVNKRLDOLRSCLSPKOHOSPALOSTD 240
Qy 241 DEVVLEGPVLPQSSRLFTLKIRCADLVRLPVMSEPLQNVVDHMANHLGVSPNRILL 300
    |||||
Db 241 DEVVLEGPVLPQSSRLFTLKIRCADLVRLPVMSEPLQNVVDHMANHLGVSPNRILL 300
Qy 301 FGESELSPTATPSTLKGVADIIDCVVLASSEETETSOELRLVQGEKHOMLEISLP 360
    |||||
Db 301 FGESELSPTATPSTLKGVADIIDCVVLASSEETETSOELRLVQGEKHOMLEISLP 360
Qy 361 DSPKVLMSHYEAMGLSGHKLSPFFDGTKLSGELPADLGLSGDLIEWG 412
    |||||
Db 361 DSPKVLMSHYEAMGLSGHKLSPFFDGTKLSGELPADLGLSGDLIEWG 412

```

```

RESULT 2
Q9GLZ9 PRELIMINARY; PRT; 408 AA.
AC Q9GLZ9:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HYPOTHETICAL 44.6 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN PARIENTAL LOBE;
RA Otsuda N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050511; BAB17279.1; -.
KW Hypothetical protein.
SQ SEQUENCE 408 AA; 44580 MW; 5E06DD066AC2F24 CRC64;

```

```

Query Match 65.8%; Score 1382; DB 6; Length 408;
Best Local Similarity 70.1%; Pred. No. 2,4e-91;
Matches 293; Conservative 31; Mismatches 78; Indels 16; Gaps 5;

```

```

Qy 1 MAEPLRGPRSGRGARARGARARCPARQSPARLIPDTVLVDVSDSEVELEV-- 58
    |||||
Db 1 MAEPLRGPRSGRGARARGARARCPARQSPARLIPDTVLVDVSDSEVELEV-- 58
Qy 59 -----ADPEVAVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDGEAPV 114
    |||||
Db 53 ARCADEVADEVAPEPVGASRDSDSDSEGADARPAEPPEVPRRRRLVLDGEAPLV 112
    |||||
Qy 115 PYVSGKVOSSNLIPDNSSLLKLCPESEDEADLTNSGSSPSSEDDALPGSPWRKKLRK 174
    |||||
Db 113 PYVSGKVKSSCLIPDDLSLKLYPPGDEVEVELADSSGLYHEGSPSP--GSPWKTKLR 171
    |||||
Qy 175 CEKEEKKMEFPPODISLPPOSSRNKSRKTEALOKLREVNKRLDOLRSCLSPKOHOS 234
    |||||
Db 172 -DKREKKTEILDNSPLSPSPRTSKRKTRALKLSEVNRKRLQDLRSCLSPPEPQ 230
    |||||
Qy 235 ALQSTDEEVVLEGPVLPQSSRLFTLKIRCADLVRLPVMSEPLQNVVDHMANHLGVSP 294
    |||||
Db 231 EQDQDEEVVLEGPVLPQSSRLFTLKIRCADLVRLPVMSEPLQNVVDHMANHLGVSP 290
    |||||
Qy 295 NRILLGESELSPTATPSTLKGVADIIDCVVLASSEETETSOELRLVQGEKHOML 354
    |||||
Db 291 SRILLFGETELSPATPRTKLGVADIIDCVVLASSEETETSRQLRVQGEKHQTL 350

```

```

Qy 355 EISLSPDPLKVLMSHYEAMGLSGHKLSPFFDGTKLSGELPADLGLSGDLIEWG 412
    |||||
Db 351 EVLSRSDPSPLKTLMSHYEAMGLSGHKLSPFFDGTKLSGELPADLGLSGDLIEWG 408

```

```

RESULT 3
O60336 PRELIMINARY; PRT; 1217 AA.
AC O60336:
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE KIAA0596 PROTEIN (FRAGMENT).
GN KIAA0596.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98290545; Pubmed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998).
DR EMBL; AB011168; BAA25522.1; -.
DR InterPro; IPR001680; .
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 1.
KW Repeat; WD repeat.
FT NON_TER
SQ SEQUENCE 1217 AA; 131097 MW; 0BC4E4C66722B8F5 CRC64;

```

```

Query Match 7.4%; Score 154.5; DB 4; Length 1217;
Best Local Similarity 23.5%; Pred. No. 0.0082;
Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps 21;

```

```

Qy 6 RGRGPRSGRGARARGARARGARCPARQSPARLIPDTVLVDVSDSEVELEVADPEV 65
    |||||
Db 455 RGRGKQDQSSPQASG-----PNRQASMSLSPGPA-----LSSDSKEDGEDTEELP 506
    |||||
Qy 66 -----VRLPAPAKPEQSDSDSEGAAE-----GPAGAPRTLVRRRRRL 106
    |||||
Db 507 ALPYLAKSTKALKASVSPALPNLSLHWEWSRAOESVGFDPAPANPSP-----RRGRW 563
    |||||
Qy 107 DPGEAPVVPVYSGKVOSSNLIPDNSSLLKLCPS--EP-EDEADLTNSGSSPSSEDDALPS 163
    |||||
Db 564 QPG-----VELSVSMIDLRLQLETLAPSLQDPQSDSLAIFSPGRKKHQALET 612
    |||||
Qy 164 G-----SPWRKKLRKC-----EKEP-----KMEFPPODISLPLOPS---SRNK 201
    |||||
Db 613 SLTSQNEKPPRPQASQPCSPYPIIRLLSQEGVFAQDLAPLIEDGIVYEPESDNPMTDT 672
    |||||
Qy 202 SRKTEALOK-----LREVNKRLQD-----LRSCLSPPKOHOSPALQSDDEVL 245
    |||||
Db 673 SEFOVQAPAKKTGLRGVYPGSRSEKSPDACSVDYSSCSLSPH--PIEDSESTPLS 730
    |||||
Qy 246 VEGFVLQSSRLFTLKIRCADLVRLPVMSEPLQNVVDHMANHLGVSPNRILLFGESE 305
    |||||
Db 731 VDG-----ISSDLEPAPDGEDEEEEGGNGP-----YGLQE 762
    |||||
Qy 306 LSPATP-----STLKGVADIIDCVVLASSEETETSOELRLVQGEKHOMLE 355
    |||||
Db 763 GSP--QTPDQEQFLKHQHEFTLASGAAPGAP--VQYVERSESISISRLFLQVOTPLREP-- 818
    |||||
Qy 356 ISLSPDPLKVLMSHYEAMGLSGHK 381
    |||||
Db 819 ---SPSSSLAMSRPAOVPOASGEQ 841

```

```

RESULT      4
035482      PRELIMINARY;      PRT;      1072 AA.

ID      035482      ID      035482;
AC      035482;
DT      01-JAN-1998 (TREMBLREL. 05, Created)
DT      01-JAN-1998 (TREMBLREL. 05, last sequence update)
DT      01-MAR-2001 (TREMBLREL. 16, last annotation update)
DE      HIGH MOLECULAR-WEIGHT NEUROFILAMENT.
GN      NF-H.
OS      Rattus norvegicus (Rat).
OC      Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX      NCBI_Taxid=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91038277; PubMed=2230956;
RA      Chin S.S., Liem R.K.;
RT      "Transfected rat high-molecular-weight neurofilament (NF-H)
RT      cossembles with vimentin in a predominantly nonphosphorylated form."
RN      J. Neurosci. 10:3714-3726(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Chin S.S.M., Liem R.K.H.;
RL      Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: TO THE INTERMEDIATE FILAMENT FAMILY.
DR      EMBL; AF031879; AAB87068.1;
DR      InterPro; IPR001664;
DR      Pfam; PF00038; filament. 1.
DR      PROSITE; PS00226; IF, 1.
KW      Colled coll.; Heptad repeat pattern; Intermediate filament.
SQ      SEQUENCE      1072 AA;      115348 MW;      899A16D574AD7BB CRC64;

```

Query Match	6.9%	Score 145.5	DB 11	Length 1072
Best Local Similarity	22.1%	Pred. No. 0.031		
Matches 95	Conservative 63	Mismatches 166	Indels 105	Gaps 20

```

0Y      4 PLKRGSPRSKGRARRKRGRCBPARRASPARL-IPDVLVDLVSDEEV-----55
Db      613 PVEAKSPAFAKSPASVAKSPG-EAKSPAFAKSPAIEVKSPTATVKSPEAKSPAIEVKSPTYVK 671
0Y      56 --LEVADVEVPVRLPAPAPEDDSDSDSGAAGAGAPRTLVRRRRRLLDPGAPV 113
Db      672 SPAEAKSPVEV-----KSPAIVKSPSEKSPAGAKSPAFAKSPVAVASPAFAKSPAFAK- 725
0Y      114 VPVYSGKVOSSLNLPIDNSSLKLCPSPDEADLTN-----SSSPSEDA---LP 162
Db      726 -PFAAKSPAFAKSPAFAKS-----PFAAKSPAFAKSPVEYKSPAFAKSPVYKGAASLAE 779
0Y      163 SGSPWRKRLRRKCEKEEKMEFPDODISPLPDS-SRNSR-----KHTPLQKLRVVK 217
Db      780 AKSPAFAKSPVKEIKPFAEVKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAK 839
0Y      218 RLQDLRSLKQKQOSPALOSTDDEVYLVGCVLPQSSRLFTLKIKRCADLVLP---VR 274
Db      840 RPADIR---SEOVYKSPAFAKSPA-----PEKEETRTKFAVAKKEEVKSPAIEVVK 886
0Y      275 MSEPIQNVDDMHANHLGVSPRILLRFGESLSPTPATPSTLKGVDIIDCVYLAASSFA 334
Db      887 AKEP-----PKV-----EEKTP-ATPKTE-----VESKKDEA 915
0Y      335 TETSELRLRVQGREKHOMLEISLSPDSPLKVLMSHYEAMGLSGHKLSEFFDGTKLSG 394
Db      916 PKDAQ-----KFAKEKEPLTE--KPKDPSGPAKKEAKKEKKAAP-----E 956
0Y      395 ELPADLGLIE 403
Db      957 ETPAKLVK 965

```

RESULT 5

ID	O9XW25		PRT; 1634 AA.
AC	O9XW25	PRELIMINARY;	
DT	01-NOV-1999	(TRMBLrel. 12, Created)	
DT	01-NOV-1999	(TRMBLrel. 12, Last sequence update)	
DT	01-MAY-2000	(TRMBLrel. 13, Last annotation update)	
DE	VIBD10A.1 PROTEIN.		
OS	VIBD10A.1.		
OC	Ctenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OX	Rhabditidae; Pelodierinae; Caenorhabdits.		
NCBI_Taxid=6239;			
[1]			
RP	SEQUENCE FROM N.A.		
RA	Harris B.;		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
RM	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=94150718; PubMed=7906398;		
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,		
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,		
RA	Crafton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,		
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,		
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,		
RA	Lichting J., Lloyd C., Momuray A., Mortimore B., O'Callaghan M.,		
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showmkeen R.,		
RA	Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,		
RA	Therry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,		
RA	Watson A., Weinstock L., Wilkinson-Spoat J., Wohlman P.,		
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RT	elegans."		
RL	Nature 368:32-38(1994).		
DR	EMBL; AL034393; CAA22308.1; -		
DR	InterPro; IPR000637; -		
DR	PROSITE; PS00354; HMG1_Y; UNKNOWN_1.		
SQ	SEQUENCE 1634 AA; 179058 MW; DC203723EAF4800 CRC64;		

Query Match	6.7%	Score 140	DB 5	Length 1634
Best Local Similarity	22.0%	Pred. No. 0.13		
Matches 88	Conservative	54	Mismatches 152	Indels 106
				Gaps 18

[illegible]

RESULT
Q9XVS4

AC 092541;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE MYELOBLAST KIAA0252 (FRAGMENT).
 GN KIAA0252.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NC NCB1;TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=NONE MARROW;
 RX MEDLINE=97191544; PubMed=9039502;
 RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
 Tanaka A., Kotani H., Miyajima N., Nomura N.
 RT Prediction of the coding sequences of unidentified human genes. VI.
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain.
 RL DNA Res. 3:321-329(1996).
 DR EMBL: D87440; BAA13382.1;
 FT NON_TER
 SQ SEQUENCE 664 AA; 75809 MW; 2C6537A98BFBADCB CRC64;

Query Match 6.2%; Score 131; DB 4; Length 664;
 Best Local Similarity 21.5%; Pred. No. 0.19;

Matches 101; Conservative 69; Mismatches 151; Indels 148; Gaps 19;

DB 37 RLIPDVLVLDVSDS--DEEVLVA-----DPEVVARLPAPAKPEDSDSDEGAA 87
 2 RVIYDSTEDSGSDNDQELSLAKRRSDSEKEBPVQ--PAASDSFSDSDDEMT- 59
 QY 88 EAGAGAPRTLVRRRRRLDPGEAPVYVSGVQSSLANLPNSSLKLCPSPEDEAD 147
 60 --FGSNKKKKKKKARKIEKKG--TWKKQANKTASGSSDKD--SSAESSAPEGEVSDS 112
 DB 148 LTNSSSPSEDDALPSGSPW-----
 113 DSNSSSSSSSDSSSEEEFHDCYGEDLMGDEDRALDEQTEKERQELFNRIEKEVL 172
 QY 168 -----RKRLR--KCKEKEKKMEFPDODISPL----PQSSRNKSR-----KH 205
 173 KRREIKKKLTKAKKKKKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 232
 DB 206 TEALQKLR-EVNRKLDLJCSLSPKQ-----HQSALDSTD 240
 233 SQAMEELKAEERKKNTAELAKKOPDKTSEYVSDDEEEDDKSEKSDRSRTSSD 292
 QY 241 DEVVLVGPVLPQSSRLFTLKICRCRADLVRLPYMSEPLQNVDMANHLGVSPNRIIL 300
 293 EEEKEEIP--PKSQ-----PVLPEEL-----NRRKLSNKKLEKRW 326
 DB 301 FGESELSPTATPSTLKG-----VADIIDCVVLASSSE--ATEPSQELRLVQ 347
 327 CHMPFPAKLTGCGVIRIGNHNHNSKPYRVAVELTGVETAKVQDGLGTFRNKGLQ 386
 QY 348 KEKQMLIEISLSPSLKVLMSHEEAMGLSGHKLSEFFDCTKLSGKEL 396
 387 DORVFLIEVSNQEFTESEFMK-WKEAMFSAQMQLPTL--DETINKREL 431
 DB

RESULT 9
 ID 09H5F9 PRELIMINARY; PRT; 704 AA.
 AC 09H5F9;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE CDNA: FLJ23471 FIS. CLONE HS111969.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCB1;TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HUMAN SMALL INTESTINE;
 RA Kawabata A., Hiki T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.
 RT "MEDO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK027124; BAF15667.1;
 SQ SEQUENCE 704 AA; 75768 MW; 858F94EBA2C1FC6 CRC64;

Query Match 6.2%; Score 130; DB 4; Length 704;
 Best Local Similarity 22.6%; Pred. No. 0.24;

Matches 100; Conservative 63; Mismatches 173; Indels 106; Gaps 20;

DB 10 PRSRGCGARRGARGRCRPARQSPARLPDVLVLDVSDSEVLVADPV----- 62
 286 PKTEAPQASPLAKPLQSSSPVILGLPSRMEPPAPL-----STSSQASALVPPAGRRNLAE 341
 QY 63 -----EYVARLPAPAKPEDSDSDEGAAEGPAG-----AP 94
 342 SSGVGRVAGSRPKPEAPMKGKSTTLTQDNSTSLQSGEDGPRGMANKPVDRSPAE 401
 QY 95 RTLVRRRRRLDP--GEAP--VVPYVSGVQSSLANLPNSSLKLCPSPE 141
 402 RTLKPKRPRALAEPRAGEAPKVSFGSVHITLTPVRDPTPRPASPGSLPARSP 461
 QY 142 PEDBADLTNSSSPSEDDALPSGSPWKKLKKCKEKEKK--MEFPDODISPL--LP 194
 462 PR--RRRLAVPASLDVCCNWLPRPEPGQEARVQSWKEEKKPHLQGRGRPLSPANVPLP 520
 QY 195 QPSSRNKSRKHTAOLKLEVRKRLDRLSCPSKQHSALDSTD-----DD--EVLV 246
 521 GEYTSVPRLHPDPLSP--EEIQRQLODIERRLDALEKGVLEKRLRAAESDDAESLAW 579
 DB 247 EGPVLPQSSRLFTLKICRCRADLVRLPYMSEPLQNVDMANHLGVSPNRIIL--FG 302
 580 DMFWLHIEKQL--LQGESLWKKSKAQRLEQGLDIEGLRLMKPELAKSLQGERR 636
 QY 303 ESELSPTATPSTLKGADIIDCVVLASSSEATTSOELRLRVQGKKRQMLIEISLSPDS 362
 637 EQEL-----LEQVYSTVND-----RSDIVSDLEDRLREG--BEDQMLRDMI---- 676
 DB 363 PLKVLMSHEEAMGLSGHKLSE 384
 677 -----EKILQKKSKF 688
 DB

RESULT 10
 ID 040363 PRELIMINARY; PRT; 635 AA.
 AC 040363;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE NM1 PROTEIN.
 GN NM1.
 OS Medicago sativa (Alfalfa).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Medicago.
 NC NCB1;TaxID=3879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96361876; PubMed=8721748;
 RA Boegre L., Jonak C., Mink M., Mesikene I., Treas J., Ha D.T.C.,
 RA Snoboda I., Plank C., Wagner E., Heberle-Bors E., Hirt H.;
 RT "Developmental and cell cycle regulation of alfalfa nucleus, a plant
 RT homolog of the yeast Nsr1 and mammalian nucleolin."
 RL Plant Cell 8:417-428(1996).
 DR EMBL: X88845; CAA61298.1; -

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu L.E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ikegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostneff A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wastaman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003841; AAF59275.1;
 DR Flybase: FBgn0033156; CG12165.
 SQ SEQUENCE 755 AA; 83536 MW; 5DA71535FE3EBCD1 CRC64;

Query Match 6.1%; Score 128.5; DB 5; Length 755;
 Best Local Similarity 24.1%; Pred. No. 0.33;

Matches 77; Conservative 48; Mismatches 122; Indels 73; Gaps 16;

QY 71 APAKPEODSDSEGAEGAGAPRT--LYRRRRRL-----DGEAVVPVYSGKQ 122
 DB 41 AEAKPR-----DSAPAKSOENSVPQOTKKRRKRLTSLAEQNEEADATANNNTS 95
 QY 123 SSLNLPDSSLLKLCPEPEDEADLTNCGSSP--SEDDALPGSGPWR--KLLKKCEKEEK 180
 DB 96 QSTRV--SNQTLAIAEDENHSTASLMPPPVPVSADTTGSGRPQRAKL-----KTEK 148
 QY 181 KMEFPDQDISLPQPSRRKSR--HTEAQLKLR--EVNKLQDLRSC-----LSP 228
 DB 149 LKE-----PSLNKMRPSPSESLVKKVESEQVSOFSQSAOLEENKLA 197
 QY 229 KQHSPALQSTDEEVVLVEGPVLQSSRLFTLKIRCRADLVRLVPRKSEPLQNVDMAN 288
 DB 198 PELPEPAETAEBOOKPEEASVTEEDVTTTLKVKAKRE--KISTEAVPPLTNVVS--TAN 254
 QY 289 HLGVPENRLLLGESELSPATPSTLKLGVADIIDCVLASSEATETSOELRLRVQK 348
 DB 255 VTIVS-----SVTTEAAR-----DDTVASNTTSTEVSKVKKKKKDV 293
 QY 349 EKHQMLEI-----SLSPDPL 364
 DB 294 ESHRPKIVERPSDLKSSPV 313

RESULT 13

Q91995 PRELIMINARY: PRT; 931 AA.
 AC Q91995;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 GN CALPSTATIN.
 OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodidae; Xenopus.
 ON NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Maricci S., Rossi C., Nardi I.,
 RT "Identification of different forms of calpastatin mRNA co-expressed in
 RT the notochord of *Xenopus* embryos."
 RL Submitted (FE8-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A271209; CAB71173.2;
 DR InterPro: IPR001259;
 DR Pfam: PF00748; Calpain_inhib. 5;
 SQ SEQUENCE 931 AA; 96825 MW; D71A35226255299E CRC64;

Query Match 6.1%; Score 128.5; DB 13; Length 931;
 Best Local Similarity 21.8%; Pred. No. 0.43;
 Matches 83; Conservative 60; Mismatches 150; Indels 87; Gaps 18;

QY 67 ARUPAPA--KEODSDSEGAEGAGAPRTLYRRRRRLDPEAP-----V 113
 DB 32 SKTPAPSGKTMMNPVDTQPKGTAGTAPAAATK-----PTAASALAKTTAPSA 79
 QY 114 VVYSGKVGSSLLNLPDSSLLKLCPEPEDEADLTNCGSSPSDDALPGSGPWRKLR 173
 DB 80 VPAATKPTAA-----PSTAPVAVTKPTEIKPSA-----GKSPKQDTRP--TPGKQKPS 128
 QY 174 KCEKEKKMEFPDQDISLPQ--PSSRNK-----SRKTEAQLKREVNKR 218
 DB 129 K-SKEEKSSDKATAPVQ--PKVTPSSAGKSAVATGFVAAGTAATVAVOSTKQDES 186
 QY 219 LQDRSCLSPK--OHQSPALQSTD--DEVVLVEGPVLQSSRLFTLKIRCRADLVRLV 273
 DB 187 KEKKAVIGDKVSSASAPATGATFALDELDTLG-----SPADIPSPK 231
 QY 274 RMSEPLQ--VDMANHLG-----VSPNRLLLGESELSPATPSTLKLGVADIIDCV 326
 DB 232 FTGPEIOTDVTTSKYVEELKRDHTIPPNRYKLIDGKGEKMAPPTPLVVAESMDDDL 291
 QY 327 -VLASSEATETSOELRLRVQK--EKHQMLEISLSPDPLKYLMSHYEAMGLSGHK- 381
 DB 292 AALSSGKSSQTCPEVKKPLEEKLEKPKSAATVATQPDLOKVKVTAETSKASIOKK 351
 QY 382 --LSFPDGTGKSGELPAD 399
 DB 352 EVKATAEASKOSTQALPSPD 371

RESULT 14

Q9VNA7 PRELIMINARY: PRT; 1167 AA.
 AC Q9VNA7;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE CG1054 PROTEIN.
 GN CG1054.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Fandel M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abt11 J.F., Agbayanti A., An H.-J., Andrews-Franknoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale B., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borckova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butkiss K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jaitai M., Kalush F., Karen G.H., Ke Z., Kemnitz J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kraavitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Moberly A.C., Morris J., Moshireli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weltschok G.M., Weltschbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003605; AAF52104.1; -
DR Flybase: FBgn0037267; CG1054.
SQ SEQUENCE 1167 AA: 126283 MW: 756CD221FBA8229D6 CRC64:

[illegible]

AC 09n00c9; 2000 (TReMBLrel. 13, Created)
 DT 01-May-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-Oct-2000 (TReMBLrel. 15, Last annotation update)
 DE CG12105 PROTEIN.
 GN CG12105.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yeandlell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayroni A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brooksstein P., Brothier P.,
 RA Butts K.C., Bussam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Meltel B., McIntosh T.C., Mcleod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.W., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefel F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirtkas R., Tector C., Turner C., Turner R., Venter J., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissensbach J.,
 RA Williams S.M., Wozniak T., Wolley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003472; AAF47521.1; -
 RC FlyBase; FBgn0035241; CG12105.
 SQ SEQUENCE 1408 AA; 156836 MW; 7BACE3A9C49574CD4 CRC64;

Query Match	6.1%	Score 127.5	DB 5	Length 1408
Best Local Similarity	20.0%	Pred. No. 0.84		
Matches	94	Conservative	50	Mismatches 142; Indels 185; Gaps 20;
Qy	15	GRGARARGARRCRCRANQSPARLLIPDVLV	-----DLVSDS-----	DEEVLVEAD 60
		:	: :	:
Db	1002	GGTR-----RASLKLAEIVGETVLVYPATAAPDPLPSKEPIIDQELVYAT		1051
Qy	61	PVEYVARLPAP-----AKPEQSDSDSEGAAGCAPAG-RTLVRRRRRLDPG		109
		: : : : :	: :	:
Db	1052	PQGVATGADIAQIATLVHESSEQOSQSSAEATKRGEDQPPDSDLATMPELKV---		1108
Qy	110	EAPVVPVYSGKVOSSLNLIIPDNSSLKLCPS-----EPF-----DEADLTN-		150
		: : : : : :	:	:
Db	1109	AGPVPPVITETVEQPHDAKNQESIVKKPKITLTKVKNISIDKPEPEOKNDSSKLSNI		1168

```
QY 151 -----SGSSPSEDDALP-----SGSPWRKKLRRKCEKEKME 183
      : : : : :
Db 1169 LRDRNKINELSSRINKLTPPKQAKPKKEAKKATTANEDSAKPEEATPAEETTAETPALE 1228
      : : : : :
QY 184 EF-PDODISPLPQSSSRNKSRRKHEALQKLEVNKRLODLRSLSPKOHOSPALOS---- 238
      : : : : :
Db 1229 EEEPEVEPEPAPKPKKIKKKV----IKRQKRRLSVGDTFFLOPEPEPEPAIPEVETI 1284
      : : : : :
QY 239 -----TDDEVVLVEGPVLQSSRLFTLIRCRADLVRLPVRMSEPLQNVVDHMANHLG 291
      : : : : :
Db 1285 ERAIAYVTDE---EDP-----KAPAEPEPEV----- 1308
      : : : : :
QY 292 VSPNRILLFGESELSPAPSTLKGVAADIIDCVLASSSEATE-----TSQ 339
      : : : : :
Db 1309 -----NELKSCLVHREYKIG-----DIVLYAERYRKTQVRKKRGVLERITSI 1351
      : : : : :
QY 340 ELRLRVQGEKEKHOMLEISLSPDSPLKVLMSHYEAMGISGHKLSFFPDGTK 390
      : : : : :
Db 1352 SYKLEIEGKE-----VPAHVSYIKKYTGKRVRF---GTK 1382
      : : : : :
```

Search completed: September 11, 2001, 08:56:05
Job time: 429 sec

This Page Blank (uspo)

~~BEST AVAILABLE COPY~~

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 ; Search time 19.29 Seconds
(without alignments)
1294.820 Million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099
1 MAEPLRGKGRPSRSGRGARR.....GKELPADLGLESGLIEWVG 412

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

A_Geneseq_0601:*

1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	18	AAW34091
2	2099	100.0	412	20	AAV08331
3	1399.5	66.7	419	20	AAV08330
4	132	6.3	586	18	AAW10423
5	124.5	5.9	518	21	AAW53322
6	124.5	5.9	671	21	AAV9426
7	124.5	5.9	671	22	AAW66175
8	123	5.9	950	20	AAV33298
9	122.5	5.8	1780	19	AAW53863
10	122.5	5.8	1780	21	AAW15380
11	121	5.8	2971	21	AAW41231

12	121	5.8	2972	22	AAW50363
13	121	5.8	3118	22	AAW50362
14	120.5	5.7	565	19	AAW61247
15	120.5	5.7	1881	21	AAW44506
16	120	5.7	669	19	AAW37483
17	118.5	5.6	508	21	AAW77945
18	118.5	5.6	846	21	AAW71057
19	118	5.6	576	20	AAW49541
20	118	5.6	576	20	AAW49545
21	118	5.6	576	20	AAW49546
22	118	5.6	576	20	AAW06477
23	118	5.6	576	21	AAW93684
24	118	5.6	576	21	AAW51932
25	118	5.6	576	21	AAW51939
26	117	5.6	979	14	AAW34783
27	115.5	5.5	1341	21	AAW85657
28	115	5.5	376	20	AAW49547
29	113.5	5.4	661	19	AAW77048
30	113.5	5.4	661	21	AAW70078
31	113	5.4	519	21	AAW44247
32	113	5.4	783	20	AAW60344
33	112.5	5.4	95	20	AAW87985
34	112.5	5.4	610	12	AAW10923
35	111.5	5.3	244	21	AAW16198
36	111.5	5.3	2286	22	AAW65635
37	111	5.3	605	20	AAW31741
38	111	5.3	897	17	AAW92751
39	111	5.3	897	19	AAW47118
40	111	5.3	897	20	AAW94406
41	110.5	5.3	1519	21	AAW85660
42	110.5	5.3	2518	21	AAW40574
43	109.5	5.2	605	21	AAW57950
44	109.5	5.2	710	20	AAW80995
45	109.5	5.2	1142	21	AAW43876

ALIGNMENTS

RESULT	1
ID	AAW34091 standard; Protein; 412 AA.
XX	
AC	AAW34091;
XX	
DT	18-MAY-1998 (first entry)
XX	
DE	Mouse NF-AT interacting protein 45.
XX	
KW	NF-AT Interacting Protein 45; NIP45; Yeast two-hybrid assay; mouse;
KW	Rel homology domain; RHD; T cell; transcription factor; cancer;
KW	interleukin-4; IL-4; development; Th1; Th2; cytokine; allergy;
KW	autoimmune disease; transplantation.
XX	
OS	Mus sp.
XX	
FT	Key
FT	Region
XX	
PN	W09739721-A2.
XX	
PD	30-OCT-1997.
XX	
PF	23-APR-1997; 97MO-US06708.
XX	
XX	25-NOV-1996; 96US-0755592.
PR	23-APR-1996; 96US-0636602.
XX	
XX	25-NOV-1996; 96US-0755584.
PA	(HARD) HARVARD COLLEGE.
XX	
XX	Human gravin polyp
XX	Human gravin prote
PI	Glimcher LH; Ho I; Hodge MR;

Human SRCAP. Homo
Human SRCAP. Homo
Streptococcus pneu
Streptococcus pneu
Mouse liver cancer
A. thaliana enviro
Human membrane tra
Human PRO201 (Nsp1
Human Nsp1 mutant
Human Nsp1 mutant
Human tumour-assoc
Amino acid sequenc
Human PRO201 prote
Human PRO201 prote
Human CENP-C anti
Human Actinus L pro
Human Nsp1 mutant
Human striated mus
Human striated mus
Human cell signal
Human normal blad
Ubiquitin-like dom
Maize protein enco
Arabidopsis thalia
Novel protein kina
Human podocalyxin-
Murine EGF recepto
Murine eps15 prote
Human tyrosine kin
Human ORFX ORF338
Human transmembran
Human guanine nucl
Amino acid sequenc

```

XX  WPI: 1997-535556/49.
DR  N-PSDB; AAT93045.
XX
PT  Production of cytokine(s) associated with Th2-type helper T cells -
PT  particularly for controlling development of Th1 and Th2 cells for
PT  treatment of allergy, autoimmune disease etc.
XX
PS  Claim 31: Fig 11; 151pp; English.
XX
CC  This is the amino acid sequence of the mouse NF-AT Interacting Protein 45
CC  (NIP45). The gene sequence was isolated by using a yeast two-hybrid
CC  detection system for proteins that interact with the NF-AT Rel homology
CC  domain (RHD). The assay used, as a "bait", a 900 bp fragment of the
CC  murine NF-ATp encoding the region spanning amino acids 228-250. NF-AT is
CC  a multisubunit transcription complex containing a cyclosporin A sensitive
CC  cytoplasmic phosphoprotein and an inducible component of the AP-1 family
CC  of transcription factors. The screen was carried out on a cDNA library
CC  prepared from the murine T cell line D10. One class of proteins, NIP45
CC  designated NIP45, bound the NF-AT-RHD region with high affinity. NIP45
CC  can be used in a claimed method to inhibit or stimulate production of
CC  NF-AT family protein, particularly interleukin-4 (IL-4) and the treated
CC  cells may be administered to control development of Th1 or Th2 cells by
CC  modulating the production of a T-helper type 2 associated cytokine.
CC  Especially the method is useful in the inhibition of Th2 in allergy,
CC  cancer or infections, and promotion of Th2 in autoimmune disease and
CC  transplantation.
XX
SQ  Sequence 412 AA:

Query Match 100.0%; Score 2099; DB 18; Length 412;
Best Local Similarity 100.0%; Pred. No. 7.4e-175;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGGRGRSGRGARRARAGRCPRAROSPAPRLIPDTYLVLDVVSDEVELEVAD 60
    |||||||
DB 1 maeplrggrprsgrgarraragrcprarqsparlldptvldvvsdsdevleavad 60

QY 61 PVEVPVRLPAPAKPEQDSDSDSGAEGPAGAPRTLVRRRRRRLDPGEAPVVPVYSGK 120
    |||||||
DB 61 pvevpvarlpapakpegdsdsdsegaagpagaprtlvrtrrrrrllldpgsapvpyvysgk 120

QY 121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSSGSPWRKLRKKCEKEE 180
    |||||||
DB 121 vqsslntlipnsslklkcpespedeadtlnsgsspsddalpsgspwrklrkckeeek 180

QY 121 vqsslntlipnsslklkcpespedeadtlnsgsspsddalpsgspwrklrkckeeek 180
    |||||||
DB 121 vqsslntlipnsslklkcpespedeadtlnsgsspsddalpsgspwrklrkckeeek 180

QY 181 KMEFPPODIIISPLPQPSRRKSKRHTALOKLREVNKRLODLRCSLSPKOHQSALSTD 240
    |||||||
DB 181 kmeefpqddisplpqpsrrkskrhtaalklrevnkrldlrlscslspkqhspalstd 240

QY 241 DEVVLVEGPVLPQSSRLFTLKIRCRADLVRLPYRMSEPLQNVVDHMANHLGVSFNRITLL 300
    |||||||
DB 241 devvlvegplvpqssrftlkircradlvrlpyrmseplqnvvdhmanhlgvsfnrlll 300

QY 301 FGSELSPTATPSTLKLGAVDIIIDCVVLASSSEATETSQELRLRVGCKEKHOMLEISLSP 360
    |||||||
DB 301 fgeselsptatpstlklgavdiidcvvllassseatetsqelrlrvvgckekhqmlleislsp 360

QY 361 DSPPLKVLMSHYEAMGLSGHKLSFFPDGTLKSGKELPADGLGSGDLEIEMWG 412
    |||||||
DB 361 dspplkvlmshyeamglsghklsffpdgtlksgekelpadlglesgdleievwg 412

RESULT 2
AAV08331
ID AAV08331 standard; Protein: 412 AA.
XX
AC AAV08331;
XX
DT 16-jUL-1999 (first entry)
XX
DE Mouse NIP45 protein.

```

```

XX  NIP45; mouse; transcription transactivator; IL-4; interleukin-4;
KW  NF-AT interacting protein-45; modulator; treatment; inflammation;
KW  autoimmune disease; HTP screening; drug testing; allergic therapy;
KW  T-cell dependent response; cytokine; diagnostic; immunosuppressant.
XX
OS  Mus musculus.
XX
PN  W09921993-A1.
XX
PD  06-MAY-1999.
XX
PF  21-OCT-1998; 98WO-GB03141.
XX
PR  24-OCT-1997; 97GB-0022388.
XX
PA  (ZENE) ZENECA LTD.
XX
PI  Liu D, Zhao J, Zhou H;
PI
DR  WPI: 1999-312964/26.
XX
PT  Polynucleotides encoding human NF-AT interacting protein (NIP45)
PS  disclosure; Fig 4; 84pp; English.
XX
CC  This invention describes a novel human NF-AT interacting protein-45
CC  NIP-45. Human NIP45 is a transcriptional trans-activation factor of the
CC  interleukin 4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be used
CC  to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or
CC  autoimmune disease. NIP45 is a good candidate target for HTP screening
CC  and/or testing system for drugs that will alleviate T-cell dependent
CC  autoimmune and allergic responses, and for cytokine-based therapies of
CC  chronic disease. Expression of NIP45 can be inhibited, and IL-4
CC  expression modulated, by administering an antisense molecule. Antibodies
CC  against NIP45 and primers can be used in diagnostic assays. Discovery of
CC  a human NIP45 is advantageous in that it provides the ability to control
CC  IL-4 transcription, which is of importance for anti-inflammation and
CC  immunosuppressant drug development.
XX
SQ  Sequence 412 AA:

Query Match 100.0%; Score 2099; DB 20; Length 412;
Best Local Similarity 100.0%; Pred. No. 7.4e-175;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGGRGRSGRGARRARAGRCPRAROSPAPRLIPDTYLVLDVVSDEVELEVAD 60
    |||||||
DB 1 maeplrggrprsgrgarraragrcprarqsparlldptvldvvsdsdevleavad 60

QY 61 PVEVPVRLPAPAKPEQDSDSDSGAEGPAGAPRTLVRRRRRRLDPGEAPVVPVYSGK 120
    |||||||
DB 61 pvevpvarlpapakpegdsdsdsegaagpagaprtlvrtrrrrrllldpgsapvpyvysgk 120

QY 121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSSGSPWRKLRKKCEKEE 180
    |||||||
DB 121 vqsslntlipnsslklkcpespedeadtlnsgsspsddalpsgspwrklrkckeeek 180

QY 121 vqsslntlipnsslklkcpespedeadtlnsgsspsddalpsgspwrklrkckeeek 180
    |||||||
DB 121 vqsslntlipnsslklkcpespedeadtlnsgsspsddalpsgspwrklrkckeeek 180

QY 181 KMEFPPODIIISPLPQPSRRKSKRHTALOKLREVNKRLODLRCSLSPKOHQSALSTD 240
    |||||||
DB 181 kmeefpqddisplpqpsrrkskrhtaalklrevnkrldlrlscslspkqhspalstd 240

QY 241 DEVVLVEGPVLPQSSRLFTLKIRCRADLVRLPYRMSEPLQNVVDHMANHLGVSFNRITLL 300
    |||||||
DB 241 devvlvegplvpqssrftlkircradlvrlpyrmseplqnvvdhmanhlgvsfnrlll 300

QY 301 FGSELSPTATPSTLKLGAVDIIIDCVVLASSSEATETSQELRLRVGCKEKHOMLEISLSP 360
    |||||||
DB 301 fgeselsptatpstlklgavdiidcvvllassseatetsqelrlrvvgckekhqmlleislsp 360

QY 361 DSPPLKVLMSHYEAMGLSGHKLSFFPDGTLKSGKELPADGLGSGDLEIEMWG 412
    |||||||

```



```
Dh 11 rarrtrarrg9g9gvsnsrths-gkcrtrgralsapflflatttttmmgyastddds 69
QY 56 LEVADPEVVALPAPAKRQSDSDSEGAAGBPACAPRTIVRRRRRLDDGCEAVVP 115
Db 70 lllktpdel-----dkysgsptlll-----tltdkhdirqpr 101
QY 116 VYSGK---VOSINLIPDN-----SSLKLCPESEDEADLTNNGSSPSSEDALPS 163
Db 102 vhrqtyhlqlhldirpeelrpfqlllstplql--geandeeg--taprtlgeeeetaas 157
QY 164 GSWRRKKLRKKCEKEKKMEFPDODISPLPQSSRNKSKRHTEALOKLREVNRLQDLR 223
Db 158 hepekk-----ekgekke-----ded-----drdddrer-----gllcvsnedsdvr 197
QY 224 SCLSS--PKOHSPALOSTDDEVDVVEGVPVLPOSSRLFTLIRCRADIVLVRPMKSEPLON 281
Db 198 pafslfparqchllrsvldq-qltmaivrlshlftlrl-llplkrlplr-----rk 250
QY 282 VDHMANHLGVSPNRILLFGESELSPPTATPSTLKLGVADIDCVVLASSSEATETSQEL 341
Db 251 aahhtalh-----dcalahlpelptfeptldlnvtenaasvad---taestda-dltppl 301
QY 342 RLKVOGKEKHOMLEISLSPSPKLKVMSHYEAMGLSG-----HKLS----- 383
Db 302 tvrvr-----halcwhrveg-glsqprgltsrtsarlsetaktlqpr 342
QY 384 FFPDGRFKLSGKELPADLGLES 404
Db 343 svfgrleldpneapdltlss 363

RESULT 5
AAB53322
ID AAB53322 standard; Protein; 518 AA.
XX
AC AAB53322:
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen protein sequence SEQ ID NO:862.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytosolic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
XX Homo sapiens.
XX OS
XX PN MO200055351-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000MO-US05883.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA; Ruben SM;
XX DR WPI: 2000-587534/55.
XX DR N-PSDB; AAC98079.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
XX Clafm 11; Page 1414-1416; 2104pp; English.
XX PS
XX AAC97991 tq AAC98763 encode the human colon cancer associated proteins,
```

```
CC called human colon cancer antigens, given in AAB53324 to AAB54006. The
CC human cancer antigens can have cytotoxic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 518 AA:
XX
Query Match 5.9%; Score 124.5; DB 21; Length 518;
Best Local Similarity 23.1%; Pred. No. 0.02;
Matches 93; Conservative 41; Mismatches 143; Indels 125; Gaps 20;
QY 8 RGPSSRCGARGARAGRCPRAROSPARLIPDTYLVLDV-----SDSD 52
Db 97 rapr-rgplgrkrkk-----kapsadsksdsgdgaekpepvamarasaasssssssdsg 151
QY 53 EBYLE-----VADPEVVALPAPAKR-----EDSDSD-----SEGAABGPACAPRT 96
Db 152 vsvkkprrgkpkpaekpripkprgkpkprrpsasssdsdsdevdrisewkrdrarrel 211
QY 97 LVRRRR-----RRL-----LDPEGAPVVPYSGKVOSSLNLIPDNSSLK 136
Db 212 earrrrgeeelrrlregekekerrrerdrgeae-----rsgsgsgdelredepyk 266
QY 137 -----LCPSEPPD--EADLTNSGSSPEDDALPSSGSPWRKRLR----- 172
Db 267 krgtrgrgrgppssdsdsepealeareakkaekpqsststeparkpqketrvrpeekqqa 326
QY 173 --KKCEKEKKMEFP--DODISPLPQSSRNKSRKHTKRALOKLR--EVNKRLO---DLRS 224
Db 327 kpykvertkrksegfsmdrivvekkkepsveek-----lqklnseikfalkvdsdpvkr 379
QY 225 CLS-----PKOHSPALOSTDDEVVLVE-----GPVLPOSSRLFTLIRCRADIVR 270
Db 380 clnaleelgltgtstqlqntdvatllkrlrykankdvmekaeevlt-----r 429
QY 271 LPVRMSEPLONVVDHMANHLGVSPNRILLFGESELSPATP 312
Db 430 lksrvtlqpkleavqkv-nkagemekekkaeklageelageeap 470

RESULT 6
AAY99426
ID AAY99426 standard; Protein; 671 AA.
XX
AC AAY99426:
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1604 (UNG785) amino acid sequence SEQ ID NO:308.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
XX Homo sapiens.
XX OS
XX PN MO200012708-A2.
XX PD 09-MAR-2000.
XX PF 01-SEP-1999; 99MO-US20111.
XX
```



```

PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100390.
PR 15-SEP-1998; 98US-0100398.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100586.
PR 16-SEP-1998; 98US-0100587.
PR 16-SEP-1998; 98US-0100561.
PR 16-SEP-1998; 98US-0100664.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 01-OCT-1998; 98US-0102965.
PR 02-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 06-OCT-1998; 98US-0103454.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.

PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0105882.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0107833.
PR 10-NOV-1998; 98US-0107775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.

PA (GETH ) GENENTECH INC.
XX
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WT:
XX WPI: 2000-237871/20.
XX N-PSDB: AAA37108.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
XX secreted PRO polypeptides, useful for screening of potential peptide or
XX small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12: Fig 174; 773bp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
XX receptor or secreted PRO polypeptides given in AAy99340 to AAy99642. The
XX transmembrane and receptor PRO proteins can be used for screening of
XX potential peptide or small molecule inhibitors of the relevant
XX receptor/ligand interactions. The polypeptides and nucleotide sequences
XX encoding them have various industrial applications, including uses as
XX pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
XX PCR primers and hybridisation probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX

```


QY	310	A-TPSTIKLGVAIIDCVIASSSEATTEISQELRLR--VOGKERHOMLEISPDSP.L	364
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	438	agvsipaeeIygmD-----aepgeapeakeIvXIketcvsgedptgdaD--IspdekV	487
QY	365	-----KVIMSHYEAMGISGKHLFFEDGT--KLSEKELPADLG--LESGLI	408
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	488	IsKpegvsvseveImssqermkvqgsplkklftstgIkkIsgkkqkrgygdseesgeht	547
QY	409	EV 410	
		:	
Db	548	qv 549	
	RESULT 10		
	AAB15380		
ID	AAB15380	standard; Protein; 1760 AA.	
AC	AAB15380;		
XX	26-JAN-2001	(first entry)	
XX	Human	gravin protein sequence.	
DE	Human	gravin protein sequence.	
KW	Human;	gravin; PKA RII binding site; myasthenia gravis;	
KM	kinase anchoring protein; CAMP dependent protein kinase.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
FT	Binding-site	265..556	
FT		/note="PKC binding site"	
FT	Binding-site	1526..1582	
FT		/note="PKA RII binding site"	
FT	Region	1537..1563	
FT		/note="PKA anchoring site"	
XX	US6090929-A.		
XX	18-JUL-2000.		
PD	19-DEC-1997;	970S-0994570.	
XX	19-DEC-1997;	970S-0994570.	
PE	19-DEC-1996;	960S-0769309.	
XX	(UYOR-) UNIV OREGON HEALTH SCI.		
PA	Klauck TM, Scott JD, Nauert JB;		
XX	WPI: 2000-523763/47.		
DR	N-PSDB: AAA74903.		
XX	Novel polynucleotides useful for detecting gravin in patients suffering		
PT	from Myasthenia gravis encodes CAMP-dependent protein kinase-binding		
PT	polypeptide and protein kinase C-binding polypeptide of gravin -		
XX	Claim 1; Column 35-45; 34pp: English.		
PS	The present sequence is the protein sequence of human gravin. Grav		
XX	in an A-kinase anchoring protein (AKAP) which is involved in the		
CC	localisation of CAMP dependent protein kinase A (PKA) via interactions		
CC	between the RII binding region and the PKA regulatory subunit R1. Grav		
CC	is also an antigen found in myasthenia gravis sufferers, and it is		
CC	thought that antibodies to it may be useful in modulating the binding of		
CC	PKA, and thus aid in the treatment of the disease. The gravin coding		
CC	sequence was isolated by first screening a human umbilical vein		
CC	endothelial cell cDNA library with serum from a myasthenia gravis		
CC	patient, and then searching a human heart cDNA library for sequences		
CC	resembling the isolated sequence. This was done because the first		
CC	sequence obtained was shown to be shorter than the full length cDNA.		
XX	Sequence 1780 AA;		
XX			

Query Match	5.8%;	Score 122.5;	DB 21,	Length 1780;
Best Local Similarity	22.0%;	Pred. No. 0.18;		
Matches	93;	Conservative	62;	Mismatches 152;
			Indels	115;
			Gaps	22;

Qy	40	PDIV-LVLDVNSDSDEEVLLEVADPVEVYVARLPAPAK---	EPQSDSDSSDEAGAPGAPR	95	
Db	192	pdtvqllltvkkdegeagaagdhqdsplgagaaskesepkqstekp	-----	240	
Qy	96	TLVRRRRRLDPEAPVVPYSGKVQSSLLNLPDNSSLLKLCPSPEPD--	EADLTNGSGS	153	
Db	241	tlkregshaelssp-----paesgqa-----	veeckeegeekqekesksae	281	
Qy	154	SPSEDDALPSSP-----WRKK--LRKCKEKE---	EKKMEFPDQDISLPDP	196	
Db	282	sptsrvtselstgftkklffftgwaqwrkktflrxpkedeveasekkekqepk	-----vdt	336	
Qy	197	SSRKSRKHTHALQKLRLRVNKRLODDLRSCLSPK---	OHSPALQSTPDEVLVEGP---	249	
Db	337	eedgaavaasekllaseqahpq-epaaeahprrlsaeekvelps-eeqvsqsgpreek	354		
Qy	250	VLPOSSRLFTLKICRADLVRLPRMSEPLQNVVDHMANHLGVSPNRILLFGSELSPT	309		
Db	395	paplatetvfdckievhegevaevaehtvsteeteeq-----	kteveec	437	
Qy	310	A--TPSTLKLGVDIIDCVLASSSEATFETSOELRLR---	VQGEKKQMLEISLSPDSL	354	
Db	438	agvsyaaeelvgmd-----aepqeaepakelyklketcvsgedprtgad--	lspkek	487	
Qy	365	-----KVLMSHYEAMGLSGHKLSPFDDT---	KLSGKRLPADLG---	LESGLD	408
Db	488	lskpregvvsevemlssqgermkvqgsplklftstgltklsgkqkxrggdeesgeht	547		
Qy	409	EV	410		
Db	548	qv	549		
RESULT 11					
ID	AAB41231	AAB41231 standard; Protein; 2971 AA.			
AC	AAB41231;				
DT	08-FEB-2001	(first entry)			
DE	Human ORF995	polypeptide sequence SEQ ID NO:1990.			
XX	Human; open reading frame; ORF95; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antihypoid; antiangiemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.				
XX	Homo sapiens.				
OS	MO200058473-A2.				
XX	05-OCT-2000.				
XX	31-MAR-2000; 2000MO-US08621.				
XX	31-MAR-1999; 99US-0127607.				

PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR N-PSDB: AAC75440.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 11: Page 1510-1517; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiproliferative; antiparasitoid; neurotropic; neuroprotective;
 CC osteoplastic; anticonvulsant; antihypertensive; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertensive; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy.
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 SQ Sequence 2971 AA:
 Query Match 5.8%; Score 121; DB 21: Length 2971;
 Best Local Similarity 21.2%; Pred. No. 0.53; Mismatches 173; Indels 166; Gaps 18;
 Matches 102; Conservative 41;
 6 RGRPRSGKRGARRARG-RC-----PRA 31
 2125 kakperpgtvtserltgaretgahntvysahqtrsttprcsparearvprpapr 2184
 32 ROSPA---RLIPDTVLVDVSDSEVLEVADEVP-PAARLPAAPAKPEQSDSDSEGAA 87
 2185 rptasapaalpalyvpyv---sapvlsapnpltlpvhllppppsqqlppspact 2241
 88 EGPGAP-----RTIVRRRRRLDGEAPVYVSGKQSSNLPLD----- 130
 2242 pppactpppactpppactclvtpssplllgp---psvpilsavtnlpgllpseelcaga 2298
 131 -----NSSILKICPSE-----PEDEDLTNSSGSSPS---EDDALP 162
 2299 laspeslelasvasetslslyppkdllyaveellpysenhlsltpsapbltleagslp 2358
 163 SGSPWRKKLKKCKEKKME-----FPDODISPLPQSSRN---KSKRHTEALQKL 212
 2359 ngq-----eqeapdsaefttlvlpgeelplcvseanglelppsaasdeplqep 2408
 213 REVKKRLQDLRSCSPKQHSALQSTDDDEVL-----VEGVLPGSSRLFTLKI 262
 2409 leadtseelkaktspkspkqgelvtaeapstssatspsspparprtr----- 2463
 263 RCRDLVRLPRMSEPLQNVVDHMANHLGVSPNRLLLFGESEI-----SPTAPSTLKL 317
 2464 rtasdelvrggtgtrpggppkylrk-----pqlvtvveekelvyrrrrgrgaastlv 2520
 318 GVADIIDCVLASSSEATETSOELRLRVQKEKHOMLEISLSPDSPKLKVLMSHYEAMGL 377

DB 2521 gvs-----etaspqspvsmgssppsl 2546
 QY 378 SG 379
 DB 2547 99 2548
 RESULT 12
 ID AAB50363 standard; protein: 2972 AA.
 XX
 AC AAB50363;
 XX
 DT 12-MAR-2001 (first entry)
 XX
 DE Human SRCAP.
 XX
 KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;
 KW cAMP regulatory element; CREB binding protein; CBP; ATPase;
 KW transcription activation; DEAD box RNA dependent helicase;
 KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.
 XX
 OS Homo sapiens.
 XX
 PN M0200073467-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 25-MAY-2000; 2000MO-US14719.
 XX
 PR 27-MAY-1999; 99US-0136620.
 PR 25-MAY-2000; 2000US-0579181.
 XX
 PA (UYSL-) UNIV SAINT LOUIS.
 XX
 PI Chrlivia J, Yaciuk P;
 XX
 DR WPI: 2001-061545/07.
 DR N-PSDB: AAC69860.
 XX
 PT Snf2 related cAMP regulatory element (CREB) binding protein (CBP)
 PT activator protein, capable of co-activating CREB binding protein,
 PT useful for modulating transcription and for affecting viral infection -
 XX
 PS Claim 1: Page 86-94; 103pp; English.
 XX
 CC The present sequence is an Snf2 related CREB (cAMP regulatory element)
 CC binding protein (CBP) activator protein (SRCAP) polypeptide. It has
 CC ATPase activity and is capable of activating transcription. SRCAP
 CC polypeptides are useful for activating transcription in a cell, for
 CC enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated
 CC activation of transcription in a cell, for treating a patient having a
 CC disease involving a function such as insufficient transcription of a
 CC gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent
 CC helicase, adenoviral DBP protein, beta-actin or a nuclear receptor
 CC affected by SRCAP protein. Compounds that modulate SRCAP function, such
 CC as antibodies, antisense molecules, polynucleotides or ribozymes, are
 CC useful for treating diseases mediated by SRCAP-activated transcription,
 CC for example, infection by adenovirus, hepatitis C virus, human
 CC immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or
 CC hepatitis B virus.
 CC
 SQ Sequence 2972 AA:
 Query Match 5.8%; Score 121; DB 22: Length 2972;
 Best Local Similarity 21.2%; Pred. No. 0.53;
 Matches 102; Conservative 41; Mismatches 173; Indels 166; Gaps 18;
 6 RGRPRSGKRGARRARG-RC-----PRA 31
 2125 kakperpgtvtserltgaretgahntvysahqtrsttprcsparearvprpapr 2184

QY	32	RSQSA	---RLIEDPVLVDIYSDSDGEVLEKADPVEY-PVARLRPAKAPKEDSDSDEGAA	87
Db	2185	rfpcsapaaipalvprv	---sarpvisapnptltllpvhlilpsppsqippcsspact	2241
QY	88	EGPAGAP	-----RTLVRRRRRRLDPGCAVPVPYSGKVGOSLNLIPD-----	130
Db	2242	pppactpppactpppactclvtspsp11lgr	---psvrlsaavtnplglipaealcag	2298
QY	131	-----NSSLIKLCPSE-----	PEDEADLTNGSSPS--EDDALP	162
Db	2299	laspeslelaavaasetsislvpkdl1pvaellpvsekhlsttpaasltleagslp	2358	
QY	163	SGSPWRKKLRKKCKEKKME	-----FPDDIISLPQSSRN-----KSRKHTPALOKL	212
Db	2359	ngq-----eqeapdsaeqtlclvlpegeelplcvseanglelppsaasdeplpe	2408	
QY	213	REVNKRLLQDLRSCSPKHOHSPALQSTHDEVVL-----	VGAPVLPQSSRLFTLKI	262
Db	2409	leadtseelcaetcltspsspekpgelvtaeaapstssatsppegsparrprt	-----	2463
QY	263	RCRADVLRLPYRMSEPLQNVVDHMANHLGVSPNRILLFGESEL	-----SPTANPSTLKL	317
Db	2464	rtseadvelrgqgtgrppqppkvrlkl--	prlvtvcekelvgqrrrgqgaastlvp	2520
QY	318	GVADIIDCVLASSSEATETISQELRLRYGCKEKKHQLLEISLPDSPKLKVMHVEEAKGL	377	
Db	2521	gvs-----	etsaspqspvsrsmgspesppi	2546
QY	378	SG	379	
Db	2547	gq	2548	
RESULT 13				
ID	AAB50362	standard; protein; 3118 AA.		
XX	AAB50362:			
AC	12-MAR-2001	(first entry)		
DT	12-MAR-2001	(first entry)		
XX	Human SRCAP.			
DE	Human SRCAP.			
XX	Human: SRCAP: Snf2 related CBP activator protein; antiviral; CREB;			
KW	CAMP regulatory element; CREB binding protein; CBP; ATPase;			
KW	transcription activation; DEAD box RNA dependent helicase;			
KW	adenoviral DBP protein; beta-actin; nuclear receptor; viral			
XX	infection.			
OS	Homo sapiens.			
XX	WO200073467-A1.			
PN	07-DEC-2000.			
PD	25-MAY-2000; 2000MO-US14719.			
XX	27-MAY-1999; 99US-0136620.			
XX	25-MAY-2000; 2000US-0579181.			
PR	(UYSL-) UNIV SAINT LOUIS.			
PA	Chrivia J, Yaciuk P;			
XX	WPI: 2001-061545/07.			
XX	DR N-PSDB: AAC89859.			
XX	Snf2 related CAMP regulatory element (CREB) binding protein (CBP)			
PT	activator protein, capable of co-activating CREB binding protein,			
XX	useful for modulating transcription and for affecting viral infection			
PS	Claim 6; Page 77-86; 103pp; English.			

Query Match	Best Local Similarity	5.8%	Score 121	DB 22	Length 3118
Matches 102	Conservative 41	Mismatches 173	Indels 166	Gaps 16	
6 RGRGRSRGRCGARRARGA	-----RC-----	PRA 31			
2271 kakapqglvserlgrnaetgqanlhpvisahqtrsttprcsparer	-----PRA 31				
32 RQSPA---RLIPDTVLVLDVSDDEVLVADPVPRV-IPARIPAPAKEDQSDSSEGA	87				
2331 rplcpasapaalpialvpvpv---sapvlsapnpltilpvhilpspppsqjprcsspact	2387				
88 EBPAGAP-----RTLYRRRRRLDPGEAFVNVYSGKQVSSINLIPD-----	130				
2388 pppactpprhbtpprpqaqclvpspslllgp---psvlsasvtnlplgltpraealcaga	2444				
131 -----NSSILKLCPSE-----PEDEADLTNMSGSSPS---EDDALP	162				
2445 laspselalasvasetslsivppkdlipvaeallipsexknlstlpaspsllleagslp	2504				
163 SGSPRRKKLRKCKEKKEMEE-----FPDODISPLPOPSRN---KSRKNTALQTL	212				
2505 ngq-----eqeapdsaeqtlcltlylpeggeelplcvseengtelppsaasdeplqep	2554				
213 REVNKRLODRSCLSPKOHQSPALQSTDEVLV-----VEGVALPOSSRLFTLKI	262				
2555 leadrtseelteakrptsppekqglvtaevaapatsstssatspspepparppr	2609				
263 RCRALVLRLPVMSRPLNLNVDMNHNHGVSPNRILLFGESEL-----SPATPSTLKL	317				
2610 rtsaavetlrgqgtgrgppqpprkvtlrl---pgrlvtveekelqvtrrtqgrgaastlvp	2666				
318 GVAIDIDCVLASSSEATETSOELRLRVQGEKKHOMLEISLSPDSPLKVLMSHYEAMGL	377				
2667 gvs-----etsaapgsapsvrsmgspsppl	2692				
378 SG 379					
2693 gq 2694					
RESULT 14					
AAM61247					
ID AAM61247 standard; Protein: 565 AA.					
XX AAM61247;					
XX 02-OCR-1998 (first entry)					
XX Streptococcus pneumoniae SPI23 protein.					
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;					
KW detection; pneumonia; otitis media; meningitis.					
XX					

Query Match 5.7%; Score 120.5; DB 21; Length 1881;
 Best Local Similarity 20.9%; Pred. No. 0.3; Mismatches 156; Indels 85; Gaps 19;
 Matches 81; Conservative 66;

```

QY 29 PRAROSP-----ARLPDTVLVDLVSDSDSEVLEADPVEV--PVARLP--APAKPE-Q 77
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 pkteespkeepkseevkptddtlpkveegkedsaepapveevgeveskpeekvaavkpesq 242
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 DSDSDSGAAGPAGARTLVRRRRRLDPGEAPVVPVYSGKYQSS--LNLIPDSSL 134
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 psdkpaeeskveqagep-vaprededkapvepekpeapeeekaveetlpkgeestlpdlkae 301
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 LKICP-----SEPDEDLTFNSGS--SPSEDDALPSSGSPWRKK 170
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 etvepkveetvngsiegpkvvetpavakqtepteebpkveqagepvaqpredaqpt-apepe 360
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 171 LRRKCEKEKKMEFFPDODISPLPQSSRNKSRKHTEALQKLREVNKRLODLRSCLSPKQ 230
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 kppevppeekavee-----tpkpedkikgigtkepvdk-selinqi-dkassvsptd 410
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 HQSPALQSTDDDEVVLVGEFVLPQSSRLFTLKIRGRADLVRLPVRMSE--PLQNVVD---- 284
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 ystasynal-----gpyletakgya-----sepvkqpevnsetnkiktaidaInv 456
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 ---HMANHLGVSPNRILLFGESELSPTATPSTLKLGVADIIDCVLASSSEATETSOEL 341
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 dkcelnltiadaktkvkehysdtrswgnlqtevtlkaekvaantd-----akqseveavekl 512
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 RLRYQGEKHHOMLEISLSPDSPLEKVLMS 369
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 tatie-----klvels---ekpiltlts 532
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: September 11, 2001, 08:49:26
 Job time: 60 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 ; Search time 13.53 Seconds
(without alignments)
626.994 Million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099
Sequence: 1 MAEPLRGGRPSRSGRGAR.....GKELPADLGESGLIEWVG 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PTUS.COMB.pep.*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	2	US-08-755-584-2
2	2099	100.0	412	3	US-09-192-611-2
3	122.5	5.8	1780	1	US-08-769-309A-5
4	122.5	5.8	1780	3	US-08-994-570-5
5	120.5	5.7	565	4	US-08-961-083-218
6	113.5	5.4	661	2	US-08-785-868-14
7	111	5.3	897	1	US-08-095-737-4
8	111	5.3	897	1	US-08-480-145-4
9	111	5.3	897	2	US-08-477-389-4
10	108.5	5.2	505	1	US-08-631-200-2
11	108.5	5.2	505	1	US-08-630-592-4
12	108.5	5.2	505	1	US-08-714-891-4
13	108.5	5.2	505	2	US-08-829-553-2
14	108.5	5.2	505	2	US-08-922-267A-2
15	108.5	5.2	505	2	US-08-936-707A-2
16	108.5	5.2	505	2	US-08-936-706A-2
17	108.5	5.2	505	3	US-09-248-203-2
18	108.5	5.2	505	3	US-09-032-365A-4
19	108.5	5.2	505	4	US-08-812-824-3
20	108.5	5.2	505	4	US-09-406-071-2
21	108	5.1	459	1	US-08-630-592-2
22	108	5.1	459	1	US-08-714-991-2
23	108	5.1	459	3	US-09-032-365A-2
24	108	5.1	673	2	US-08-455-073A-6
25	108	5.1	1001	4	US-09-060-410-2
26	107	5.1	521	2	US-08-721-684C-2
27	107	5.1	521	2	US-09-005-970-2

28	107	5.1	521	4	US-09-407-715-2	Sequence 2, Appli
29	107	5.1	570	3	US-08-826-246-2	Sequence 2, Appli
30	107	5.1	570	3	US-08-944-495-2	Sequence 2, Appli
31	107	5.1	570	4	US-09-126-640-7	Sequence 7, Appli
32	107	5.1	570	4	US-08-925-588-2	Sequence 2, Appli
33	106.5	5.1	723	2	US-08-548-159-5	Sequence 5, Appli
34	106.5	5.1	1142	2	US-08-993-118-7	Sequence 7, Appli
35	106.5	5.1	1142	2	US-08-845-528C-7	Sequence 7, Appli
36	104.5	5.0	861	1	US-08-484-105-18	Sequence 18, Appli
37	104.5	5.0	861	1	US-08-484-106-18	Sequence 18, Appli
38	104	5.0	126	2	US-08-853-974-1	Sequence 1, Appli
39	104	5.0	126	4	US-08-172-988-1	Sequence 1, Appli
40	104	5.0	3248	1	US-08-353-700-1	Sequence 1, Appli
41	104	5.0	3248	5	PCT-US95-16216-1	Sequence 1, Appli
42	103.5	4.9	586	2	US-08-630-822A-70	Sequence 70, Appli
43	103.5	4.9	586	2	US-09-005-069-70	Sequence 70, Appli
44	103.5	4.9	704	1	US-08-188-582-5	Sequence 5, Appli
45	103.5	4.9	704	1	US-08-646-715-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-755-584-2
: Sequence 2, Application US/08755584
: Patent No. 5858711
: GENERAL INFORMATION:
: APPLICANT: Glincher, Laurie H.
: APPLICANT: Hodge, Martin R.
: TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
: TITLE OF INVENTION: OF USE THEREFOR
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/755,584
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Giulio A. DeConti, Jr.
: REGISTRATION NUMBER: 31,503
: REFERENCE/DOCKET NUMBER: HUI-026
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 412 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-755-584-2

Query Match 100.0%; Score 2099; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 1 6e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAEPLRGGRPSRSGRGARARAGRGRCPRARQSPARLIPDTVLVDLVSDSEVLEVAD 60

|||||
Db 1 MAEPLRGGRPRSGRGARRARGARCPARARQSPARLIDPTVLYVLDVSDDEVLEVAD 60
QY 61 PVEVPVRLPAPAKPEDSDSDSGAEGPAGAPRTLYRRRRRLDPGEPVVPVYSGK 120
Db 61 PVEVPVRLPAPAKPEDSDSDSGAEGPAGAPRTLYRRRRRLDPGEPVVPVYSGK 120
QY 121 VOSSLNIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
Db 121 VOSSLNIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
QY 181 KMEFPDQDISPLPQPSRRKSRKHTALOKLREVNRLDRLSCLSPKQHSFALOSTD 240
Db 181 KMEFPDQDISPLPQPSRRKSRKHTALOKLREVNRLDRLSCLSPKQHSFALOSTD 240
QY 241 DEVYLVGPVLPOSSRLFTLKIRCRADLVRLPVMSSEPLQNVVDHMANHGLGVSPNRILL 300
Db 241 DEVYLVGPVLPOSSRLFTLKIRCRADLVRLPVMSSEPLQNVVDHMANHGLGVSPNRILL 300
QY 301 FGESELSPTATPSTLKLGVADIIDCVVLASSSEATETSOELRLRVQGEKQHMLEISLSP 360
Db 301 FGESELSPTATPSTLKLGVADIIDCVVLASSSEATETSOELRLRVQGEKQHMLEISLSP 360
QY 361 DSPKLVMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGESGDLIEWWG 412
Db 361 DSPKLVMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGESGDLIEWWG 412

RESULT 2

US-09-192-611-2
; Sequence 2, Application US/09192611
; Patent No. 6090561
; GENERAL INFORMATION:
; APPLICANT: Glimcher, Laurie H.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,611
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUT-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-192-611-2

Query Match 100.0%; Score 2099; DB 3; Length 412;
Best Local Similarity 100.0%; Pred. No. 1,6e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGGRPRSGRGARRARGARCPARARQSPARLIDPTVLYVLDVSDDEVLEVAD 60
Db 1 MAEPLRGGRPRSGRGARRARGARCPARARQSPARLIDPTVLYVLDVSDDEVLEVAD 60
QY 61 PVEVPVRLPAPAKPEDSDSDSGAEGPAGAPRTLYRRRRRLDPGEPVVPVYSGK 120
Db 61 PVEVPVRLPAPAKPEDSDSDSGAEGPAGAPRTLYRRRRRLDPGEPVVPVYSGK 120
QY 121 VOSSLNIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
Db 121 VOSSLNIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
QY 181 KMEFPDQDISPLPQPSRRKSRKHTALOKLREVNRLDRLSCLSPKQHSFALOSTD 240
Db 181 KMEFPDQDISPLPQPSRRKSRKHTALOKLREVNRLDRLSCLSPKQHSFALOSTD 240
QY 241 DEVYLVGPVLPOSSRLFTLKIRCRADLVRLPVMSSEPLQNVVDHMANHGLGVSPNRILL 300
Db 241 DEVYLVGPVLPOSSRLFTLKIRCRADLVRLPVMSSEPLQNVVDHMANHGLGVSPNRILL 300
QY 301 FGESELSPTATPSTLKLGVADIIDCVVLASSSEATETSOELRLRVQGEKQHMLEISLSP 360
Db 301 FGESELSPTATPSTLKLGVADIIDCVVLASSSEATETSOELRLRVQGEKQHMLEISLSP 360
QY 361 DSPKLVMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGESGDLIEWWG 412
Db 361 DSPKLVMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGESGDLIEWWG 412

RESULT 3

US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauack, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein


```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-218

Query Match      5.7%; Score 120.5; DB 4; Length 565;
Best Local Similarity 20.9%; Pred. No. 0.019;
Matches 81; Conservative 66; Mismatches 156; Indels 85; Gaps 19;

QY 29 PRAROSP-----ARLIDTVLYVSDSEVLEADPVEV--PVARLP--AAKPE-Q 77
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 34 PKEESKEPEKSEVKPTDPTLPRKEGKDSAPAPVEEVEGGESEKPEEKAVKPEEQ 93
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 DSDSDSGAAEGPAGARTLVRRRRRLDPGEAPVVPVYSGKVQSS---LNLIPDNSSL 134
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 94 PSOKPAESVVEQAGEP-VAPREDEKAPVEKEPEAPEEKAVVEETPKQESTPDTKAE 152
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 LKICP-----SEPDADLTNCSG--SPSEDDALPSSGPPMRKK 170
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 EYEPKKEEYFNQSIQPKVETPAVEKQTEPEEKVEQAGEPAPRDEQAPT-APEPE 211
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 171 LRKCKEKKMEFPDIPDISLPQSSRNKSRKHTALQKLRVNRRLDLRSCLSPKQ 230
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 KQPEVPEEKAVEE-----TPKPEDKIKIGTKEPVDE-SELNNOI-DKASSVSPTD 261
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 HOSFALOSTDEEVLVGPVLPQSSRLFTLKIRGRADLVRLPVMSSE--PLQNVVD--- 284
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 262 YSTASYAL-----GPVLETFKGVYA-----SEPVKQPEVNSSETNKLKTAIDALNV 307
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 ---HMANHLGVSPNRIILFGESELSPTAPPSYLKIGVADIIDCVLASSSEATSEQL 341
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 DKTELNTIADAKTKVKEHYSDRSMQMLQTEVTFAEKVAAANTD---AKQSEVNEAVEKL 363
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 RLRRQGEKQHOMLEISLSPSPKLVMS 369
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 TATLE-----KLVELS---EKPIILLTS 383
   | : || : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-795-868-14
; Sequence 14, Application US/08795868
; Patent No. 5846773
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,868
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 661 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-795-868-14

Query Match      5.4%; Score 113.5; DB 2; Length 661;
Best Local Similarity 20.0%; Pred. No. 0.096;
Matches 103; Conservative 55; Mismatches 131; Indels 227; Gaps 25;

QY 4 PLRGR---GPRSRG---RGARRRGAR----- 25
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 26 PARGHPIVAPPSHSGACAGHHRRREAREVYVRLPGHRSRPQTPLSAEGRLMA 85
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 26 -GRCPRAROSPARIPTVLYVSDSE--EYVEADPVEVPV-----ARLPAPAK 74
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 LGSNPLRVKAGSRL-----DKLQFEERRSLIENSPPAPLRPVPLRKASLEQPK 139
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 PEODSDSDSGAAEGPAGARTLVRRRR-----RLDPGEAPVVPVYSGKVQSS---GK 120
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 SERGAPWGTGASQSEELRAGVSAERRRLFOQKASLDETRQGRSPASDLELRAGQLGR 129
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 VOSSLN---LIPDNSSL---LKICPSEPE--DEADLTNCSGSPSEDDAL----- 161
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 IRSTSEELVRSHELSRLATQRAPSPREPGEPEPLFSRPSTPSTRAYSVAQAQPSPS 259
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 ---PSSGPPMRKKLRRKC-----EKKKMEFP-----DDIISLPQPS----- 197
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 AEKPGDEGRPRSRGPAGRTPEGPGQOEVRRRDQFPLTRSRAIOECRSVPPPADPPE 319
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 198 SRNKS---RKHTALQKLR-----EVNKRLO----- 220
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 ARTKAPRGKRREPPAQAVRLPMATPGLGAAVPTLEKNRAGPEAKRLRRRGPEEDPW 379
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 ---DLRSCLSPKQHQ-----SPALQSTDEEVL-----VEPVLPQSSRLFTLKIRCRAD 267
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 GPMWRGARGAQGRARAPTSPELESSDSDSVSAGEEPLAPV----- 422
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 268 LVRLPVMSSEPLQNVYVDMANHLGVSPNRIILFGESELSPTATPSTLKIGVADIIDCV 327
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 423 -----FELPLQNVV-----VAP-----GADVLLACTI 444
   | : || : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-08-095-737-4
; Sequence 4, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: Difiore, Pier P
; APPLICANT: FazioII, Francesca
```


Db 541 HAEQNNLESEPTHOESSVSSPEIAPSDV--TDESAVTVAGNEKYTPRFDDKHSKEE 598
QY 396 LPADLGLESGDLIE 409
Db 599 DP--FNWESSSLTD 610

RESULT 9
US-08-477-389-4
; Sequence 4, Application US/08477389
; Patent No. 5872219
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,389
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/095,737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH60,001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-389-4

Query Match 5.3%; Score 111; DB 2; Length 897;
Best Local Similarity 21.2%; Pred. No. 0.24;
Matches 92; Conservative 79; Mismatches 153; Indels 110; Gaps 24;

QY 61 PVEVVARLPL-----APA-KPEOD-----SDSDSEGAAG-----PAGAPRTL 97
Db 202 PMSIPPALVPPSKRTWTVAEAKATDEIFLTKDKMDGYSGLEVREFFLTGLPSAL 261
QY 98 VRRRRRLDPGEAPVVPVYS-----GRVOS-----SLNLIPDNSSLK-----LC 138
Db 262 LAH-----INSLCDTKGCGKLSKDFALAHFLI--NQKLKIGIDPPHSLT 304
QY 139 PS--EPDEADLTN--SGSSPSED-----DALPSSPPRRKILRKCKEKKMEFPD 187
Db 305 PEMIPDRSLCKNITGSSPVADFSAIKELDTLNNEIIVLDIREKNVNEODLKEKEPTVK 364
QY 188 QDISPLPQSSRNKSRKTEALOKLREVNRKLDLRSCLSPKHQ-----SPALQSTPDE 242
Db 365 QRTSEV--QDLQDEVQESINLQRLQAKQOVQVDELLGELDEQKQDLQELQAEVRRKCAEE 422

QY 243 VLV---EGFVLPOSSRLFTLK---IRCRADLVLRPMRSEPLQNV-----VDHMANHL 290
Db 423 AOLISLKAETTSQESQISSYEELLKARELSRLQOETALQEESEVSGAQLPEPQHL 482
QY 291 GVSPNRILLFGESELSPAT-----PSTKLGVADITIDCVLASSSEATETSO 339
Db 483 QESQOETISSMOMRLKMDLETDDNNQSNWSSSPQSVLVNGATDY--CSLSSTSSSETNFMK 540
QY 340 --ELRLRVQGEKHEKQMLEISPD--SPKVLMSHYEGAMQLSGH--KLSPFDQTKSGKE 395
Db 541 HAEQNNLESEPTHOESSVSSPEIAPSDV--TDESAVTVAGNEKYTPRFDDKHSKEE 598
QY 396 LPADLGLESGDLIE 409
Db 599 DP--FNWESSSLTD 610

RESULT 10
US-08-631-200-2
; Sequence 2, Application US/08631200
; Patent No. 5646040
; GENERAL INFORMATION:
; APPLICANT: Kleyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/631,200
; FILING DATE: 12-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-631-200-2

Query Match 5.2%; Score 108.5; DB 1; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY 3 EPLR-----GRGPRSGGARGARARGRCRRAROSPARLIPDVIVLDVSDSEEVY 56
Db 44 EPLMVQANADGR--PRSR-----RAQSEDA--PLVESYLSSSGSTSY 83
QY 57 EVADPVEVVARLPAPAKPRQSDSDSEGAEE--GPAGAPRTLVRRRRLDLPGEAPVY 114
Db 84 QVEDADSIASVOGATRRPARASAKSKGAAGGOGGAPRRKKKKKNGTSGPA----- 138
QY 115 PVYSGVQSSSLNLIPDNS-----SLKLCPSEPEDEADLTNSGSSPSDDALPSSGPMR 168

Db 139 -----TLAEDKSEAQPVQILTVGOSDHDKDGAGTAAGG-----AQPSSGODLR 182
QY 169 KKLKKCKEKKKEEPPDDDISPLPPSSRNKSRKTEALQKIREVKNKRLQDLRSLSP 228
Db 183 ATMQRKGISSMSFDEDEDED-----ENSSSSQLNSNTPRSSATSATRSKSIREAASAPSP 236
QY 229 KOHOSPALOSTDDEVVLVEGPVLPPQSSRLFTLKIRCRAD 267
Db 237 AAPEPP-----VDIEVODLEEFALRPAPQGITIKCRITRD 271

RESULT 11

US-08-630-592-4
; Sequence 4, Application US/08630592
; Patent No. 5770432
; GENERAL INFORMATION:
; APPLICANT: Nishina, Patsy
; APPLICANT: No. 5770432entrauth, Konrad
; APPLICANT: Nagert, Juergen
; APPLICANT: No. 5770432th, Michael
; TITLE OF INVENTION: Obesity Associated Genes
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 941114187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MSDOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,592
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J.
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: A59504/BIR/PJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 7811989
; TELEFAX: (415) 3983249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-630-592-4

Query Match 5.2%; Score 108.5; DB 1; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY 3 EPLR-----GRGPRSRGCGARRARGRCPPAROSPRLIPDYLVLDVSDSEVL 56
Db 44 EPLMVQANADGR-PRSR-----RAQSEBOA---PLVESYLSGSGSTSY 83
QY 57 EVADPVEVPYARLPAPKPEODSDSEGAEE--GPAAPFTIVRRRRRLLDGEAPV 114
Db 84 QVQEADSIASVOLGATPPAPASAKSKGAASGOGAPRKEKKGHKGTSGA-----138
QY 115 PYVSGKVOSSLNLPDMS-----SLKLCPSEPEADLTNSGSSPEDDALPSSGSPWR 168
Db 139 -----TLAEDKSEAQPVQILTVGOSDHDKDGAGTAAGG-----AQPSSGODLR 182

QY 169 KKLKKCKEKKKEEPPDDDISPLPPSSRNKSRKTEALQKIREVKNKRLQDLRSLSP 228
Db 183 ATMQRKGISSMSFDEDEDED-----ENSSSSQLNSNTPRSSATSATRSKSIREAASAPSP 236
QY 229 KOHOSPALOSTDDEVVLVEGPVLPPQSSRLFTLKIRCRAD 267
Db 237 AAPEPP-----VDIEVODLEEFALRPAPQGITIKCRITRD 271

RESULT 12

US-08-714-991-4
; Sequence 4, Application US/08714991
; Patent No. 5776762
; GENERAL INFORMATION:
; APPLICANT: NORTH, Michael
; APPLICANT: NISHINA, Patsy
; APPLICANT: No. 5776762en-Trauth, Konrad
; APPLICANT: NAGERT, Juergen
; TITLE OF INVENTION: OBESITY ASSOCIATED GENES
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,991
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: A-59504-1/PJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-8771
; TELEFAX: 415-494-8700
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-714-991-4

Query Match 5.2%; Score 108.5; DB 1; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY 3 EPLR-----GRGPRSRGCGARRARGRCPPAROSPRLIPDYLVLDVSDSEVL 56
Db 44 EPLMVQANADGR-PRSR-----RAQSEBOA---PLVESYLSGSGSTSY 83
QY 57 EVADPVEVPYARLPAPKPEODSDSEGAEE--GPAAPFTIVRRRRRLLDGEAPV 114
Db 84 QVQEADSIASVOLGATPPAPASAKSKGAASGOGAPRKEKKGHKGTSGA-----138
QY 115 PYVSGKVOSSLNLPDMS-----SLKLCPSEPEADLTNSGSSPEDDALPSSGSPWR 168
Db 139 -----TLAEDKSEAQPVQILTVGOSDHDKDGAGTAAGG-----AQPSSGODLR 182
QY 169 KKLKKCKEKKKEEPPDDDISPLPPSSRNKSRKTEALQKIREVKNKRLQDLRSLSP 228
Db 183 ATMQRKGISSMSFDEDEDED-----ENSSSSQLNSNTPRSSATSATRSKSIREAASAPSP 236

QY 229 KQHSPALQSTDEVLVEGPVLPOSSRLFTLKIRCRAD 267
: | | | | : : : : :
Db 237 AAEPP-----VDIEVODLEEFALRPAPOGITIKCRITRD 271

RESULT 13
US-08-829-553-2
; Sequence 2, Application US/08829553
; Patent No. 5817762
; GENERAL INFORMATION:
; APPLICANT: Kieyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennle & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/829,553
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/631,200
; FILING DATE: 12-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-829-553-2

Query Match 5.2%; Score 108.5; DB 2; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;
QY 3 EPLR-----GRGPRSRGGGARRARGRCPPARQSPARLLPDIYLVVSDSDSEVL 56
: | | | | : | | | | : : : : :
Db 44 EPLMVQANADGR-PRSR-----RARQSEBOA---PLVESYLLSSGSTSY 83
: | | | | : | | | | : | | | | : : : : :
QY 57 EVADPVEVPVAPARLPAPAKPPRODSDSDEGAEE--GPAGAPRTLVRRRRRRLDPGEAPVY 114
: | | | | : | | | | : | | | | : | | | | : : : : :
Db 84 QVEADSIASVQIGATRPAPASAKSKGAASGGGAGPRKKRKHGTSQPA----- 138
: | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
QY 115 PYVSGKVQSSLLNIPDINS-----SLKLCPSEPEDEADLTNSGSSPSEDDALPSGSPWR 168
: | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
Db 139 -----TLAEDKSEAGVPVQILITVGQSDHKDAGETAAGC-----AOPSGODLR 182
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
QY 169 KLRKKCEKKEKKMEFPDODISPLPQSSSRNKSRRKHTALQKLREYVKRLQDLRSCISP 228
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
Db 183 ATKQKGISSMSFDEDEDED-----ENSSSSQULNSNTRPSSATSRKKSIREAASAPSP 236
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
QY 229 KQHSPALQSTDEVLVEGPVLPOSSRLFTLKIRCRAD 267

Db 237 AAEPP-----VDIEVODLEEFALRPAPOGITIKCRITRD 271
: | | | | : | | | | : : : : :
Db 237 AAEPP-----VDIEVODLEEFALRPAPOGITIKCRITRD 271

RESULT 14
US-08-922-267A-2
; Sequence 2, Application US/08922267A
; Patent No. 5861239
; GENERAL INFORMATION:
; APPLICANT: Kieyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennle & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,267A
; FILING DATE: 2-SEP-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/829,553
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/631,200
; FILING DATE: 12-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-922-267A-2

Query Match 5.2%; Score 108.5; DB 2; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;
QY 3 EPLR-----GRGPRSRGGGARRARGRCPPARQSPARLLPDIYLVVSDSDSEVL 56
: | | | | : | | | | : : : : :
Db 44 EPLMVQANADGR-PRSR-----RARQSEBOA---PLVESYLLSSGSTSY 83
: | | | | : | | | | : | | | | : | | | | : : : : :
QY 57 EVADPVEVPVAPARLPAPAKPPRODSDSDEGAEE--GPAGAPRTLVRRRRRRLDPGEAPVY 114
: | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
Db 84 QVEADSIASVQIGATRPAPASAKSKGAASGGGAGPRKKRKHGTSQPA----- 138
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
QY 115 PYVSGKVQSSLLNIPDINS-----SLKLCPSEPEDEADLTNSGSSPSEDDALPSGSPWR 168
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
Db 139 -----TLAEDKSEAGVPVQILITVGQSDHKDAGETAAGC-----AOPSGODLR 182
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
QY 169 KLRKKCEKKEKKMEFPDODISPLPQSSSRNKSRRKHTALQKLREYVKRLQDLRSCISP 228
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : : : : :
QY 229 KQHSPALQSTDEVLVEGPVLPOSSRLFTLKIRCRAD 267

